

A strategy tinged with chauvinism

There may be more to the new NIH strategy document than a mere projection of the past into the future, but it is a great misfortune that the document is so inward-looking.

THE US National Institutes of Health (NIH) are, as a corporate unit, the largest and most successful research organization in the world. There can be no dispute about that. It is not merely that the annual budget of NIH amounts to \$9 billion in round numbers, but that NIH has consistently held that the most effective route to the improvement of public health is the deepening of the understanding of human biology. With the encouragement of a sympathetic (sometimes self-interested) Congress, NIH is not merely dominant in clinical research, but has also been the chief source of support for the basic research that has made the United States a gusher of new knowledge in the field. It is a remarkable success, not simply an illustration of what money can accomplish.

So why should such an organization need what it calls a "strategic plan"? One answer is that Dr Bernadine Healy, NIH director for the past 18 months, has set her heart on having one. Her ample justification is that even (or, perhaps, especially) large organizations land in trouble without a clear understanding of what they are about. She has been resolute in telling NIH's scientist-dependants that if they do not help her set priorities, somebody less starry-eyed will do it for them. Dutifully, she has also done her best to involve the research community in the evolution of the plan; more than 2,000 researchers have shown up at the four 'town meetings' organized to discuss earlier drafts of a document that began at 800 pages, which is now down to half that length and which may have shrunk to 100 pages when staff editors have finished with it over the summer. Yet the research community is unconvinced.

That is not surprising. Part of the trouble is that the evolving plan is almost exclusively a statement of goals, saying little about the means by which they will be achieved. (To each set of objectives is attached a set of criteria, often banal, by which success is to be judged.) That is why the American Society of Cell Biologists protested early on that the plan made no commitment to a target for the number of grants to individual investigators. For their trouble, the cell biologists have won a declaration that "investigator-initiated research is at the heart of scientific inquiry" — but no target number. (But there is an intelligent proposal that beginning researchers should be able to compete for modest research grants movable from one institution to another and of being used as platforms for more ambitious applications.) The truth is

that the relationship between NIH and the academic research community is strictly symbiotic; it will be interesting to see how quickly the devotion NIH now promises to 'optimal cost' research management leads to the long-overdue examination of the balance between intramural and extramural research.

In any case, the readers of the strategic plan who should be worried are not cell biologists but more traditional medical researchers. Superficially at least, the plan is a recipe for modernity at NIH. Baldly, it says that molecular medicine, biotechnology, molecular immunology, structural biology and "cellular and integrative biology" are the areas within which the greatest benefits will be won in the immediate future. (Curiously, neuroscience in its own right is given only scant attention.) And the latest draft rightly muses on the great need of funds there will be when the world's human genome projects bear fruit, setting off a scramble for an understanding of the mechanisms of familiar (not necessarily inherited) diseases. The plan's concern with technology transfer, and with the mechanisms of NIH's collaboration with industry as well as academic institutions, is in the circumstances understandable.

The industrial connotations are plainly part of the strategy behind the strategic plan. The next step is that the plan will be mulled over by NIH's political bosses at the Department of Health and Human Services and eventually by the White House. If it survives intact, in the face of suspicions that it is yet another agency plan to burst the federal budget, it will no doubt become the centrepiece of an appeal to the Congress for continued generosity. Although the buzzword 'competitiveness' was dropped from the draft at the consultation two weeks ago at Bethesda out of deference to researchers' other-worldliness, there is enough in the plan to sustain a claim that what NIH does in the next decade will create a powerful biomedical industry. To ensure that the lesson is not hidden, the present working title is nothing less than "Advantage America!"

That beguiling slogan is, sadly, the strategic plan's most obvious fault. Although the document acknowledges that the human genome project is an international enterprise, it is otherwise inward-looking. To the extent that NIH is supported by the federal government of the United States, it is right that its prime concern should be the health of the US population. (The plan announces strengthened programmes of research in the health of women and of minori-



ties.) But even though NIH research may be the largest single component of the world's biomedical research, the implicit assumption that it is both a self-sufficient and a self-contained enterprise is a serious mistake.

At the least, it will give offence elsewhere, in places where the loss of talented people to NIH seems to be endemic, for example. More seriously, especially when NIH boasts in its plan of having one of the "most aggressive" of technology transfer operations in the US government, this chauvinistic view of NIH's role promises endless repetitions of the last year's tactless decision to patent partial genome sequences. Nobody will complain if NIH's endeavours help to create a still more prosperous biomedical industry in the United States; given the multinational character of the enterprises concerned, others will also benefit. But the full exploitation of the new knowledge now accumulating will outstrip even NIH's resources, however they may be augmented by the US Congress. NIH would have been more winning of opinion outside Washington if it had referred occasionally to its need of partnerships with other smaller, but not negligible, biomedical research enterprises. □

Why not plutonium?

The time has come to think again, and constructively, about plutonium as a nuclear fuel.

WHO, other than some Japanese, says that Japan is not imaginative? While most other industrialized countries have given up building nuclear power stations and have even resolved to dismantle those already completed, Japan has been increasing nuclear generation. Now it proposes building a number of fast reactors so as make some use of the fissile material locked up in nuclear weapons in the republics of the former Soviet Union (see story on page 362). Even environmentalists should be delighted that these weapons surplus to requirements (and the provisions of bilateral arms control agreements with the United States) will not finish up as so many holes in the grounds or caverns deeper in the Earth's crust. But why are the Japanese proposing that their reactors should not fully use the neutrons generated in the fission process, but instead be operated in a way that simply converts the initial charge of fissile material into relatively short-lived radioisotopes?

In principle, a tonne of natural uranium is the energetic equivalent of one million tonnes of crude oil. Half a century ago, when people first dreamed that there would be a nuclear industry, it quickly became plain that winning the full energetic potential of uranium would require that isotopes such as uranium-238 should first be converted into fissile material, most naturally the isotopes of plutonium (of which plutonium-239 is most suitable for making weapons). So, by the 1950s, every nuclear programme spelled out a prospect of how fissile material would be extracted from spent thermal reactor fuel and used for driving fast reactors designed not simply for the production of energy

but also for the conversion of natural (or even depleted) uranium into further fissile material, plutonium as it happens.

Japan's decision not to follow that route in its proposal for disposing of unwanted Russian bombs stems partly from its wish not to be thought by other states to be stockpiling plutonium, which has acquired a bad name comparable only with that of the chemical called 'dioxin'. (There are economic reasons as well; secure storage costs money and, more important, stockpiling plutonium leaves economic resources unused.) But is it not an offence against reason that restraints of this kind should determine the ways in which economic resources are now used? Given the threat of global warming and the likelihood that countries with the technical competence of Japan will operate fast reactors safely, would it not be preferable to dust off the careful studies carried out in the 1970s of how plutonium could be safely stored in internationally secure repositories against the time when its use as a source of energy would be realistic? □

Recession without end

Governments, especially the British, should be more realistic about the ending of the recession.

THE British government is going through a bad patch because it keeps promising that the long recession is about to end, only then to discover that it does not. It would be well advised to promise something else, perhaps a celebration to mark the turn of the next century, which is only seven and a half years away. The explanation of this state of affairs is well catalogued. The clamant need for the industrialized West to invest in Central Europe and the Commonwealth of Independent States, not to mention the really poor countries of the world, comes at a time when people are still burdened by debts assumed in the heady 1980s and also fearful for their jobs. In the circumstances, it is remarkable that the economies of two-thirds of the members of the Organisation for Economic Cooperation and Development (OECD) are still growing, if more modestly than in the recent past. Britain is one of those that is not.

But that is not just bad luck, nor does it imply that the consequences will be unimportant. The high rate of commercial bankruptcy will radically alter the pattern of industrial activity and will thus shape the pattern of the economy in which growth resumes when the recession eventually ends. For those affected, it is small comfort that the process is simply a further adjustment in the process by which Adam Smith's division of labour between interdependent mercantile states would eventually make a prosperous place of Europe as a whole. How to reconcile to that prospect the host now clamouring for British exit from the European Monetary System? By making sure that those still able to sell goods on world markets, even when £1 is worth DM2.95, are adequately patted on the back. □

NIH, under fire, freezes grant for conference on genetics and crime

Washington. Funding has been frozen for a forthcoming conference on the links between genetic disposition and crime after minority groups complained to its sponsor, the US National Institutes of Health (NIH), that the conference might be a veiled attack on African-Americans.

The critics said that the conference — called “Genetic Factors in Crime: Findings, Uses and Implications” — was the latest manifestation of a federal “violence initiative” intended to find and treat biological causes of crime and violence. Because African-Americans make up 46 per cent of the US prison population, critics say that research correlating genetic traits to crime is likely to impugn African-Americans in particular.

Last week, the NIH director, Bernadine Healy, decided to delay a grant worth more than \$100,000 awarded to the University of Maryland Institute for Philosophy and Public Policy, sponsor of the conference. Healy learned of the controversy earlier this month from staff in the minority affairs and equal employment offices at NIH, which received dozens of calls after panellists on two talk shows on the cable network Black Entertainment Television attacked the conference as racist.

An NIH spokeswoman says that the agency hopes to release the funds once the conference agenda is revised to address these concerns. The conference is scheduled for 9–11 October in College Park, Maryland.

The conference was peer-reviewed and approved last year by the ethical, legal and social implications branch of NIH’s National Center for Human Genome Research. David Wasserman, organizer of the conference, calls the decision to freeze funding “an abdication of the integrity of the review process” stemming from a “misinterpretation of the purpose of the conference”.

Wasserman concedes that the brochure advertising the conference may have oversimplified some of the issues to be discussed. It states that “[g]enetic research holds out the prospect of identifying individuals who may be predisposed to certain kinds of criminal conduct, of isolating environmental features which trigger those predispositions, and of treating some predispositions with drugs and unintrusive therapies”.

However, Wasserman points out that genome administrators had no objections to the brochure when they read it, and he says that the NIH’s decision to freeze the grant after receiving objections from minority groups is a form of “coercion”.

One of the most vocal critics of the conference — and any research on the biological roots of crime — is Peter Breggin, director of the Center for the Study of Psychiatry in Bethesda, Maryland. In the current edition of the centre’s newsletter, Breggin links the conference to controversial remarks made earlier this year by Frederick Goodwin, then director of the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA). Speaking to a federal advisory committee, Goodwin pointed to similarities between the behaviour of inner-city minority youth and monkeys subjected to environmental stress (see *Nature* 356, 6; 1992).

Goodwin’s remarks drew strong criticism from the Congressional Black Caucus and other legislators, and within a few weeks his political bosses announced that Goodwin was stepping down to become director of the National Institute for Mental Health, one of three institutes within ADAMHA. At ADAMHA, Goodwin laid the groundwork for a federal “violence initiative” to investigate the biological roots of crime.

Research investigating the links between genetics and crime is a “fake subject”, Breggin says, with no data that show a correlation between genetic factors and crime. “A conference [held] in the absence of such studies is racist”, he adds.

Upon hearing the complaints, Healy decided to invite to NIH three of the scientists who reviewed the proposal for the conference to discuss their recommendations. Although one of the reviewers could not remember the proposal, another confessed to having second thoughts about the conference. In the absence of a “ringing endorsement” from the reviewers, says Joanna Schneider, an NIH spokeswoman, “we didn’t have the feeling that we could defend [the conference]” against public criticism.

Rather than cancelling the grant, NIH asked Wasserman to form a multiracial advisory panel to examine the questions being raised. Although the panel has yet to meet, Wasserman expects that it will recommend that the scope of the conference be broadened to cover race relations, the social responsibilities of scientists and the need to consider the role of poverty and other environmental factors in understanding crime.

Although Wasserman says that the conference may be given a new title, the original concept will not change. “We don’t want people to think that we ever assumed the existence of genetic factors”, he says. “But we also have no intention of renaming the conference ‘The Myth of Genetic Factors in Crime’”.

Christopher Anderson

Society asks Brazil’s president to resign

São Paulo. Brazilian science, in the midst of a terrible funding crisis, has also become embroiled in a political fight with the country’s president, Fernando Collor de Mello.

The federal secretary for Science and Technology, Hélio Jaguaribe, traded verbal punches earlier this month with the president of the Brazilian Society for the Progress of Science (SBPC), Ennio Candotti, after Candotti called on Collor to resign to avoid possible impeachment. Brazil’s Congress is investigating the activities of Paulo

César Farias, a businessman who handled the financing of Collor’s presidential campaign in 1989 and who is now accused of using his position for personal gain. “He is turning it into the Brazilian Society for the Politicization of Science”, says Jaguaribe about Candotti.

Joining the secretary in his attack on Candotti is José Goldemberg, a former science secretary and current Minister of Education, who is one of Collor’s closest political advisers. Ironically, Goldemberg played an active role in national politics as SBPC president.

Goldemberg asked the society’s main council to disown Candotti’s statement, but instead the council backed its president. In retaliation, Jaguaribe threatened to withdraw funding for the society’s various programmes, including its annual meeting, held two weeks ago in São Paulo. Jaguaribe retreated after he realized that it would repudiate the judgement of his secretariat.

The dispute takes place as funding from the government’s leading source of basic academic research has shrunk by more than 40 per cent from 1990 to 1991, with an even sharper decline expected this year. The secretariat’s main funding agency, the National Council for Scientific and Technological Development, has frozen new requests for grants and scholarships because it has been unable to distribute most of its funds for the current year.

Ricardo Bonalume

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Candotti in fight.

British say Pasteur Institute slighted their help on AIDS test

Paris. Researchers at the Institute of Cancer Research (ICR) in London say that their patience is at an end after waiting eight years for the Institut Pasteur to acknowledge their contribution to the development of the Pasteur's AIDS blood test.

The situation has not prevented the ICR's Robin Weiss, a central figure in the disagreement, from maintaining good relationships with several researchers at the Pasteur. Just last week, in fact, Weiss completed a

be filed within a month. In July 1984, Weiss authorized the Pasteur to use the infected CEM line.

The Pasteur used the CEM line for the development of its test and, thanks to a peculiarity of this line — its lack of class II HLA antigens — there were fewer false positives with the Pasteur than with the US blood test. Weiss holds a patent on a process used by the Wellcome Foundation to manufacture a blood test for the HIV virus that differs from that developed jointly by researchers at the Pasteur and the US National Institutes of Health (NIH).

It was two years before the Pasteur replied. On 12 February 1986, Montagnier said in a letter to Weiss that "the Pasteur Institute, and particularly myself, is willing to find with you an agreement that will recognize your contribution". In March the president of the institute's board, François Jacob, wrote to Weiss to confirm that sentiment. And in June, Montagnier and

Danielle Berneman, of the Pasteur's patent office, met ICR administrators in London. The last communication was a letter, dated 1 August 1986, from Berneman to ICR that "I shall get in touch with you again in September to discuss in full detail our agreement."

In 1986, the Pasteur was locked in a legal dispute with NIH over differing claims to ownership of the royalties to the blood test and was not eager to have its 1984 offer to ICR made public. NIH scientists led by Robert Gallo had, in fact, accomplished the same feat for which Pasteur had proposed co-ownership with ICR, namely, mass-producing the AIDS virus on a T-cell line. Pasteur officials acknowledged then that ICR "had a case" for being recognized but asked it to remain patient.

Pasteur officials say that Weiss's CEM cell subline was contaminated by mycoplasma and that Montagnier had to reinfect it, thus reducing the value of ICR's contribution to the blood test. But Pasteur officials told ICR two years later that they still felt an agreement was appropriate.

The experience has soured Weiss on further collaboration with the institute. "I have little appetite to enter into new negotiations... because the Institut Pasteur has not honoured a previous cooperative agreement", he wrote in March 1992, replying to a query for a cooperative agreement involving another retrovirus, HTLV-I.

Bernard Seytre

Industry slows research gains in Portugal

Lisbon. The gap between the quality of research at universities and that done in industrial laboratories is hampering the growth of Portuguese science, according to a report published earlier this month by the Organisation for Economic Cooperation and Development (OECD). The report also cites the lack of mobility among Portuguese researchers as an obstacle to developing centres of excellence.

The OECD report praised the marked improvement of Portuguese science, particularly since 1986, and says that some of it is "excellent by international standards". That improvement is due in part to the creation in the past decade of 50 nonprofit research associations, run jointly by universities and private companies, that attempt to bypass the bureaucracy that can strangle government-owned laboratories. But the report warns that the changes have occurred "in an unbalanced way", without clear priorities for stimulating economic development.

Industry, considered the weakest link in science and technology research in Portugal, employs only 25 per cent of the country's 6,500 researchers. Renowned for its lack of innovation, the private sector still focuses largely on low-technology products and prefers to buy expertise from abroad rather than make use of local talent. Portuguese companies have filed for only four European patents in the past two years.

Three years ago, with the help of the European Communities, the government launched a four-year, ECU304 million (US\$410 million) programme to support some 2,000 doctoral students and to strengthen the country's research infrastructure. The effort is expected to raise the country's annual investment in science beyond its current level of 55 billion escudos (US\$440 million).

But this programme, half of which is funded by the European Communities, may create its own problems if Portugal fails to increase the proportion of its resources — 0.8 per cent of the country's gross national product — going into research. Although the government says it expects most scientific graduates to take jobs with industry, the OECD report predicts that there will not be enough jobs to go around.

Despite the close links between Portuguese scientists and foreign researchers and the large number of Portuguese who are trained abroad, Portugal does not suffer a brain drain. Most researchers return to their original university after working abroad. Those who remain, according to the report, rarely move to another part of the country or switch universities. **José Victor Malheiros**



Weiss, left, and Montagnier . . . still good friends.

three-month sabbatical at the institute. Although ICR has asked for a share of the French royalties from the AIDS blood test (in 1991, the Pasteur received about US\$1 million from its licensees), Weiss says that the issue is "one of honour".

At issue are the results of a collaboration begun when Pasteur researchers could not grow enough AIDS virus (then called LAV) to test blood on a large scale, let alone to produce test kits for sale. In February 1984, the Pasteur sent a sample of LAV to ICR with the specific, but not exclusive, goal of "infecting cell lines". Researchers at ICR successfully infected a T-cell line known as CEM and in May sent it to the Pasteur.

The French were grateful. In a letter dated 15 June to Weiss, then director of ICR, Jean-Claude Chermann, signing for Luc Montagnier, wrote that "the line can now be considered as a good virus producer. . . . It is therefore of potential use for medical and industrial applications, and the Institut Pasteur Fondation (IPF) will be interested for its application to AIDS and related diseases and to share the property with you and your institute. Enclosed is an agreement prepared by our legal department."

The proposed agreement says that the Pasteur and ICR are co-owners of the infected T-cell subline and the process of continuous production of retroviruses, that the Pasteur "will pay a royalty" to ICR if the strain or process is used for industrial purposes and that a joint patent application will

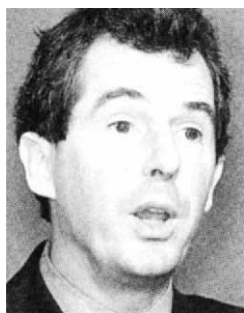
British report real decline in spending on research

London. British spending on research and development declined in value between 1989 and 1990, according to the government's annual review of research spending in the United Kingdom (*R&D 1992*, HMSO £28.00). The two years are the most recent for which figures are available.

The decline in real spending from £9.239 billion to £9.187 billion (measured in 1985 prices) chiefly reflects a decline in the government's own spending on research. There has been a steady decline in the real cost of defence research and development, which peaked (in cash terms) in 1985, and which has declined in every subsequent year except 1989 (which explains the recorded drop between 1989 and 1990).

More worrying is what seems to be a stagnation of industrial spending on intramural research, which seems to have settled at 1.5 per cent of gross domestic product (GDP) since 1986 (when it amounted to 1.6 per cent of GDP).

One curious feature of the statistics now published is that while the government's contribution to industrial research spending has declined from 30 per cent in 1967 to 17 per cent in 1990, industry's own share of the cost has remained at about 68 per cent. The difference is made up by the category 'overseas', including such sources as the European Commission and defence contracts with overseas organizations.



William Waldegrave

The annual review's comparison of British research spending with that overseas is again unflattering. The British share of GDP spent on civil science declined from 0.63 per cent in 1985 to 0.51 per cent in 1990. Of the G-7 states (with Sweden replacing Canada) cited in the annual review, only Japan (notoriously mean with government grants) and the United States spent a smaller proportion of GDP on civil science (0.43 per cent and 0.44 per cent respectively in 1990).

The decline of spending on research and development seems to have been matched by a decline of employment on these functions. The annual review says that the number of technical people employed by government has shrunk steadily from 57,000 in 1983 to an estimated 35,000 in the current year. A further 1,500 government scientists are expected to be lost in the next two years, mostly from the research councils and defence research establishments.

Industrial employment has done little to take up the slack, although the number of scientists and engineers had grown from 77,000 to 85,000 in 1989 before falling by 5,000 in the following year. Support staffs, especially technicians, have fallen sharply over the period, with the result that total technical employment in industry fell from 185,000 in 1983 to 165,000 in 1990.

In the nature of annual reviews (this was required in 1981 by the House of Lords Select Committee on Science and Technology), the text offers little in the way of promises. But on this occasion, the introductory pages draw attention to the appointment of William Waldegrave as the first minister of science in Britain for 30 years. The text comes as close as it could to offering the prospect that things will be different without promising that they will be. **John Maddox**

Japan prepares spending increase

Tokyo. An advisory body to the Japanese Ministry of Education, Science and Culture last week recommended that the budget for competitive research grants for university researchers should be raised "at an early stage" by 50 per cent, to ¥100 billion (US\$800 million) a year.

The Japanese cabinet is expected to welcome the suggestion: earlier this year, it announced its intention to double government spending on research to 1 per cent of the country's gross national product (GNP). The move should provide a significant boost in funds for Japan's cash-starved university researchers.

The ministry's competitive research grant

budget, which stands at ¥64.6 billion, has been growing in recent years at an annual rate of 5–10 per cent, well above inflation. Even so, researchers complain that many worthwhile research proposals are rejected because of insufficient funds. At the same time, the ministry has held down the budget for the fixed maintenance grants of a few million yen that go automatically to each university research department as part of its policy to reward centres of research excellence. The new recommendation conforms to the ministry's efforts to put a larger portion of the research grants budget on a competitive basis.

David Swinbanks

Wellcome Trust to double spending after sale of shares

London. The Wellcome Trust plans to double its spending on medical research next year after receiving £2.16 billion (US\$4.1 billion) from the sale of part of its holdings in the pharmaceutical company Wellcome PLC. The sale, completed last Monday, 27 July, will enable the charity to become the largest single source of medical research funding in Britain.

Officials of the trust have begun meetings with other interested parties on how to allocate that additional money. The sale will double the medical charity's spending on research from £110 million this year to £220 million in 1993–94.

In 1986, the charity sold a quarter of the company's stock; since then, the success of the AIDS drug AZT has boosted the share price considerably. The current sale of around 30 per cent of the company's stock still leaves the trust holding 42 per cent. The money from the sale will be reinvested in higher-yielding stocks, with some of the increased dividends spent on research.

The consultations — with government, the Medical Research Council (MRC), medical schools and universities, and the Association of Medical Research Charities — are intended to identify priorities and ensure that the funding from each of these bodies remains complementary. Although no details have been announced, the Wellcome Trust's director, Bridget Ogilvie, has spoken of the need for capital equipment and new buildings. Large projects are currently under way at the Royal Postgraduate Medical School and University College, London.

Last week, the trust confirmed that it would spend £50 million in the next five years on a genetic research centre that will, among other things, house John Sulston's team working on *Caenorhabditis elegans* (see *Nature* 357, 99; 1992). A site for the new centre, named after Cambridge biochemist Frederick Sanger and funded jointly with the MRC, has not been chosen.

Ogilvie also said that the trust is concerned about the insecure conditions under which researchers are employed. "Too many scientists are on 'soft money', short-term contracts without any certainty of renewal", she says. "This must be wrong."

Working parties will be set up to review the trust's funding of equipment in universities and to find ways to increase its long-standing commitment to tropical medicine, both indigenous diseases and those in which expression differs according to environment. The trust is also interested in bolstering such areas as the training of clinicians in research, psychiatry and research into vision, veterinary medicine, mathematical biology and taxonomy.

Ian Mundell

Serbian researchers labour under tightening collar of UN sanctions

Belgrade. Serbian scientists are reeling from the effects of international sanctions imposed on Serbia to stop the war in Bosnia-Herzegovina while taking part in a national movement to topple the president of Yugoslavia, Slobodan Milošević. Barely 100 km from the front lines, researchers in and around Belgrade are cut off from the rest of the world and must work without supplies, journals and salaries.

On 1 June, the United Nations imposed cultural and economic sanctions on Serbia in an attempt, so far unsuccessful, to stop the violence in the republics of the former Yugoslavia. Within days, more than 800 scientists from all over Serbia had formed an action group, Scientists for the Salvation of Serbia, to demand the government's resignation and the installation of an interim constitutional assembly that would hold general elections. The protest soon spread to the University of Belgrade, where nearly half of its 70,000 students occupied the university's main building for 26 days. All examinations were suspended until late August. Similar protests were launched in Serbia's other major cities.

In the meantime, researchers continue to work as best they can. They know that their conditions will deteriorate further and that Serbian science will most probably grind to a halt unless the demands of the United Nations are met.

Once the most open of the communist countries, Yugoslavia has a history of international scientific collaboration, most of which has been interrupted. At Belgrade's Institute of Physics, which employs 248 people, ten joint Serbia-US programmes, each worth around US\$14,000, have been suspended, along with exchanges funded by France. A \$65,000 computer, paid for before the embargo, sits in Vienna waiting to be delivered.

At the Institute of Nuclear Sciences, Serbia's biggest national interdisciplinary laboratory at Vinča, 13 collaborations each from the United States and Germany have been stopped, as well as others with Britain, Spain, France and Switzerland. Two major projects from the European Communities have been

put on ice. The biochemistry division is working at one-fifth capacity as its reserves of chemicals dry up. Research director Bogomir Dimitrijević, whose team of 40 researchers last year proved that mismatch DNA repair occurs in plants as well as animals, predicts that all experiments will stop within three months if supplies are not replaced.

The project least affected by the sanctions is the institute's ambitious TESLA cyclotron, planned by the European high-energy physics laboratory (CERN) in Ge-

neva as a user facility for the Balkans. A \$4-million housing complex has already been built for the TESLA accelerator, scheduled for completion in 1995. The cost of the \$16-million project has been provided mostly from international sources, including the International Atomic Energy Authority (IAEA) in Vienna.

Although CERN says it will not discuss any changes in the TESLA programme until next spring, project leader Nebojša Nešković is worried about its fate. Seven members of the development team are working at CERN, but because Serbian banks are international outcasts it is almost impossible to pay them. The accelerator will be used in both medicine and experimental physics, opening up such fields as heavy-ion nuclear physics and the physics of exotic nuclei as well as providing radiation therapy to treat surface cancers and radioisotopes for diagnostic and therapeutic purposes. "When the war stops we hope to supply radioisotopes to all Balkan countries, including Croatia", says Nešković.

Researchers fear that prolonged sanctions will destroy Serbian science. The next generation of scientists are seeking sanctuary abroad in huge numbers — 200,000

since last autumn. Yugoslavia has always encouraged research students to train abroad — it is estimated that more than half of Serbia's physicists live and work in other countries — but fewer and fewer are now coming back. For example, more than 20 per cent of the staff at the Institute for Nuclear Sciences, which employs 920 people, have left in the last two or three months.

Communications with other countries are becoming more difficult. Most libraries have received few journals this year; the Institute of Physics has received only 40 of its 180 subscriptions at irregular intervals, and none since the embargo. Travel has been sharply curtailed: airports are closed, car owners are allowed only 20 litres per month of petrol, currency has become worthless and invitations to international meetings have, in any case, mostly been cancelled. Two weeks after sanctions were imposed,

BITNET electronic mail lines were disconnected, stranding scientists now too poor to pay long-distance telephone bills. When supplies of fax paper run out, their isolation from the rest of the world will be complete.

The sanctions affect all aspects of life. Salaries have not kept up with a rate of inflation of around 100 per cent per month; in the past year the value of researchers' pay has diminished by 90 per cent. The embargo also means that a burst water main can close research for an extended period because the replacement parts are not manufactured in Serbia.

Researchers are divided on the merits of the embargo. The majority complain that it does not recognize the actions of groups such as theirs that are fighting the government in their own way. "As scientists we can only protest, we can't take up guns and kill people", says Dragon Vučković, head of chemistry at the Institute of Nuclear Sciences. But others say it is necessary to stop the bloodshed. "Our situation is terrible, but it is only a symptom — the war is orders of magnitude worse", says Iztok Čadež of the institute's department of atomic and subatomic physics.

Vladimir Glišin, director of Belgrade's



Work continues at Belgrade's Institute of Physics, left, while graffiti convey the wishes of many, including those awaiting the new accelerator, lower right.



Institute of Molecular Genetics, even sees a benefit to science — if the embargo does not continue for too long. A group from the institute, now working in the United States, was among the first to develop an automated DNA chip that improves sequencing, and Glišin believes that a major shakeup may open the door to the type of evaluation and peer review common in the West.

Researchers are doing what they can to blunt the impact of the embargo: Glišin admits to having driven his ice-buckets to Trieste, Italy, via Hungary and Austria, to sustain a collaboration. But the best hope for Serbian science is that the peaceful protests of researchers, students and others may help to end the political stalemate and the UN sanctions.

Alison Abbott

German states offer to trade power for money

Munich. Germany's 16 states want the federal government to pick up a higher percentage of the cost of research, leaving them with more money for academic reform.

Research funding is shared equally by the federal and regional governments. All other aspects of higher education are the sole responsibility of the individual states. But the recent crisis of university overcrowding has forced the states to look for money from other sources to implement urgent reforms (see *Nature* 357, 349; 1992).

For the first time in more than 25 years, the states are willing to reduce their power in Germany's overall research planning in exchange for the cash they need. Nevertheless, their preference is for the federal government to reallocate tax revenues to give them enough money to cope with both higher education and research.

That proposal has been criticized since it was made at last month's annual meeting of regional presidents. The federal ministry for education and research says that it cannot afford to spend more; earlier this month, its own request for an annual budget of DM2 billion (US\$1.35 billion) was rejected by the finance ministry. Its budget will instead be frozen at DM1.6 billion.

The Deutsche Forschung Gemeinschaft, the federal agency that administers research grants to universities throughout Germany, has warned that changing the balance of local and federal power over funding could damage research. But the alternative — changing the way in which tax revenues are distributed — seems unlikely in Germany's current economic crisis.

The state governments hope to discuss their proposal at an unprecedented meeting that the German chancellor, Helmut Kohl, agreed two months ago to hold. State officials are still waiting for Kohl to announce a date.

Alison Abbott

Japanese report shows cracks in predicting earthquakes

Tokyo. Japan's earthquake prediction scientists have admitted for the first time that there is a huge gap between their public optimism and the realities of research.

Any mention of shortcomings in a 30-year effort to find signs of imminent earthquakes is tucked away in a 72-page review of Japan's earthquake prediction programme released on 17 July by the Ministry of Education, Science and Culture (MESC) that is devoted to a detailed account of the observations under way. But there are also a few lines of criticism, after outsiders were used for the first time to assess the project.

Japan has invested about a billion dollars in equipment for earthquake prediction. About 500 researchers work in the programme, receiving almost ¥7 billion (US\$56 million) a year in government support in addition to their salaries. This makes the Japanese programme the longest, largest and most comprehensive effort in the world.

But the programme cannot point to any real successes, and there have been some embarrassing failures. For example, in June and July 1989, earthquake swarms around the city of Ito on the Izu peninsula south of Tokyo were for weeks attributed to plate tectonic movement. Although local fishermen had reported discoloration of the sea and dead fish on the surface, scientists did not realize that a major submarine volcanic eruption was imminent until a few hours before it occurred. Even so, most Japanese remain convinced that the next major earthquake in the seismically active region south of Tokyo can be predicted, and researchers encourage that belief.

The present review was carried out by an eight-man working group of the earthquake prediction subcommittee of the Geodetic Council of Japan, an advisory body to MESC. All the members have worked on earthquake prediction for years.

However, as a result of severe criticism of the programme by scientists such as Robert Geller of Tokyo University, the work of the committee was reviewed by six outsiders. They included Masuo Ida, a former MESC official and a vocal critic of the programme; Saburo Nagakura, president of the Graduate University for Advanced Studies and an influential figure within the education ministry; and the heads of the volcanological and seismological societies of Japan (see *Nature* 356, 464; 1992).

Although the comments of the outside reviewers have not been made public, scientists who attended hearings in May say they

were "very critical" of the programme. One Tokyo University professor says they were "perhaps too critical". Nagakura is said to have questioned both the quality of past reviews and the premise that earthquake prediction is possible.

The criticisms are not immediately apparent in the revised review released on 17 July. But page 70 of the 72-page report admits that although great improvements have been made in the network for observing anomalous phenomena associated with earthquakes such as changes in ground water levels and radon levels and fluctuations in the Earth's magnetic and electric fields, "there are many cases in which such anomalous observations cannot be objectively identified as earthquake precursors". In the absence of any successful observations, says Geller, this is the closest that Japan's earthquake prediction community has come to agreeing with its colleagues elsewhere that earthquake prediction is not possible.

On the next page, the report offers even more self-criticism. "Society's expectations for earthquake prediction are high", it states, "but we cannot deny that there is a considerable gap between society's expectations and the present state of earthquake prediction".

The report concludes by saying that the whole system of review of the programme is now "under study". In the autumn, a committee will be formed to plan the next five-year programme, which begins in 1994. All eyes are on MESC and the Geodetic Council to see if outsiders, perhaps even non-Japanese, will be included.

But critics are sceptical as to whether drastic changes will be made in the programme. Outside researchers complain that it takes too long to obtain data from Japan's observation networks, and that the only way to get data quickly is to become coauthors with members of the earthquake prediction community. Last month, for example, the Earthquake Research Institute of Tokyo University released hypocentre and travel time data recorded in 1986 on magnetic tapes.

The report calls for greater interaction and exchange of data with Earth scientists involved in basic research, but adds that the volume of data makes this impossible at present. Geller, however, points out that researchers from the United States, France and Britain have long enjoyed comparatively quick access to such data. Reform cannot begin, he says, until there is a fundamental change in the attitude of Japan's earthquake prediction researchers.

David Swinbanks

Japan plans furnace reactor to consume Soviet plutonium

Tokyo. Japan hopes to use its investment in the development of fast-breeder reactors to dispose of plutonium left by the dismantling of nuclear weapons in the former Soviet Union.

Japan is trying to interest Western nations and Russia in a plan to burn up in a giant 'plutonium furnace' reactor one hundred or so tons of plutonium taken from dismantled Soviet weapons. Japan would use its expertise in fast-breeder technology, which uses plutonium as fuel, to design the reactor and would ask Western industrial nations to help build it in Russia.

Government officials have been quietly talking about the idea since April to representatives from other countries. They persuaded the seven industrial nations (Japan, United States, United Kingdom, Canada, France, Germany, and Italy) meeting earlier this month in Munich to include it in a statement calling for the group to help Russia dismantle its nuclear weapons.

Japan is one of the few advanced nations that have continued to pursue the development of fast-breeder reactors since the technology was developed in the 1950s. The government and industry have invested ¥600 billion (US\$4.8 billion) in the construction of a prototype breeder reactor to begin operations in the next few months. But it will be decades before Japan is ready to build commercial breeder reactors.

The proposed plutonium furnace reactor differs from a breeder reactor in that it does not produce more plutonium, but their fundamental designs are similar. The breeder reactor has a core of plutonium/uranium

fuel surrounded by a blanket of uranium. Bombarding the uranium blanket with neutrons from the core creates new plutonium. The furnace reactor would lack the uranium blanket and thus not generate plutonium. However, both have a liquid-sodium cooling system that transports the vast amounts of heat produced by the plutonium/uranium core to its generators.

The Power Reactor and Nuclear Fuel Development Corporation (PNC), which is affiliated with the Science and Technology Agency (STA), has begun preliminary design work, and STA officials plan to make a request next month for money in fiscal year 1993. Quite apart from applying the technology to dispose of waste, PNC would like to develop plutonium furnace reactors that can consume the vast amounts of plutonium that Japan will have by early next century.

In the autumn, Japan will begin shipping plutonium, derived from spent Japanese nuclear fuel, back from reprocessing plants in Europe. By 2010, the government estimates it will have some 80 tons (see *Nature* 352, 7; 1991).

Japan is considering furnaces reactors of 800 MW and 1,300 MW; two of the larger reactors could use up all the plutonium from former Soviet nuclear weapons in about 30 years and produce enough electricity to power a medium-sized city. But Toichi Sakata of STA's nuclear fuel division estimates that it will be "at least ten years" before a reactor is operating. The United States and most nuclear nations are concerned that the disposal of nuclear weapons materials threatens wglobal security.

David Swinbanks

University officials dismiss quick payoff from research funds

Washington. A US presidential science advisory panel writing a report on the health of academic research was warned last week that science is hurt by exaggerated claims about the economic benefits of research.

Six years after a similar federal report extolled the payoff from basic research, politicians and the public assume that having a research centre in their neighbourhood will bolster local prosperity, said Robert Rosenzweig, president of the Association of American Universities, a consortium of 58 research universities in the United States and Canada. Such attitudes, he added, have also indirectly contributed to the growing number of research projects funded in the home districts of legislators without having received any formal peer review.

"It is now widely believed...that science and technology are the keys to local and regional economic development in the short run", Rosenzweig testified, adding that the half-a-billion dollars that the US Congress spent last year on so-called porkbarrel research projects is a monument to this belief. As research grows and funding stagnates, he said, "the political forces for wider distribution are so strong that program resources will be spread ever thinner, making concentrations of quality harder to sustain".

The hearing on 24 July, one of six held around the country in recent weeks, gave members of the academic community a chance to talk about their university's relationship with the federal government. It was convened by the President's Council of Advisors on Science and Technology (PCAST), which will use the testimony to prepare a report due out by the end of the year.

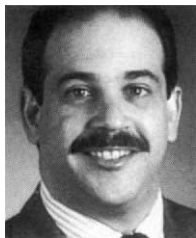
A PCAST member, Harold Shapiro, president of Princeton University and moderator for the day-long hearing, agrees with Rosenzweig that the dollar-value of academic research to the surrounding community has been exaggerated. "Some universities oversold the benefits [of research] to local constituencies," he says, adding that such boosterism may now "turn on us" when those laboratories fail to come up quickly with commercial products.

PCAST's predecessor, the White House Science Council, produced a similar document in 1986 for the federal Office of Science and Technology Policy that recommended bigger budgets and less regulation of science. But PCAST members believe that the climate has changed dramatically as the recession and the enormous federal deficit jeopardize the continued growth of research budgets.

Traci Watson

NEWS IN BRIEF

Herndon, Virginia. Spurred by new laws encouraging government scientists to



Reid Adler of NIH

learn the latest techniques.

The job can be controversial, as one of the association's founders — Reid Adler, the director of the technology transfer office at the National Institutes of Health (NIH) — learned after he persuaded his agency to file applications for patents on gene fragments identified last year by NIH

scientists. Although the Bush administration has backed NIH, the case illustrates the continuing debate that surrounds the commercialization of government research. The new association held its first meeting earlier this month.

C.A.

Washington. Mission controllers have turned off the last operating transmitter aboard Magellan, in orbit around Venus, after its degrading signal became close to unusable. The probe, which has mapped more than 97 per cent of the planet since it reached it in 1990, first had trouble with its main data transmitter more than a year ago. Its backup transmitter will be turned on briefly in September to map an area in Venus's southern latitudes, amounting to about 1.5 per cent of the planet's surface, that has never been mapped before.

C.A.

The philosopher for science

Hermann Bondi

Karl Popper's ideas have been the touchstone for judging science during much of this century. Here, in acknowledgement of his 90th birthday on 28 July, is a celebration of the man and his works.

SCIENTISTS are a very heterogeneous group of people, as befits a profession that values originality above all. Thus it comes about that only a minority of working scientists question what the nature of science is, or try to explore the demarcation between science and other human endeavours. But among this set of scientists, the name of one student of the philosophy of science brings out the deepest feelings of appreciation and, indeed, gratitude: Karl Popper. His seminal work of well over half a century ago (*Die Logik der Forschung*) is still the basis of how we think about our subject, is still the touchstone of whether one's ideas are scientifically meaningful, or just a jumble of ingenious and perhaps satisfying thoughts.

Popper's teaching in this and in later works on related themes stresses that science is a creative subject. The new ideas, the working hypotheses, the novel experimental arrangements, these are all the result of intellectual jumps and of original thinking, and not of logical deduction or inference. Thus the generation of science cannot be mechanized. There is no possibility of defining a 'scientific method', a prescription that anybody can follow and 'make science'. Scientists have to be people of flesh and blood, of passion and drive, of daring and of courage. (It is unfortunate that in some places the absurd image of the cold, passionless scientist is still propagated.) Directly or indirectly, this point of Popper's has affected our value judgements of scientists and their work. Thus in all those many cases where judging panels have given high marks to originality, Popper's influence shines through.

His basic idea is of theories having to be vulnerable to empirical disproof, with the more rigid and therefore more at-risk theory to be viewed as preferable to the more flexible (or more flabby). This view has profoundly influenced me and many others. The whole relation between experiment and observation on the one hand, and theoretical constructions on the other, is generally seen nowadays in the light of Popper's analysis. The notion of the crucial experiment to disprove a theory antedates Popper, but the appreciation that this is the principal function of experiment and observation we owe to him.

Here again Popper pointed out many years ago what is still not understood

by large sections of the population: the true relationship between science and technology. It is unfortunately widely thought that science is primary and technology derivative. In fact, advances in science frequently arise from a novel and more searching experimental method which has become available through technological progress which thus generates new science. As Popper put it so appealingly: "We advance like a person walking through a swamp, first painfully

The Black Cloud puts this so well when a non-scientist says of scientists: "I cannot understand what makes them tick. They are always wrong and they always go on." This very popperian sentiment inspires us all through our trials and tribulations. Popper has made it clear that we should be proud so to be described.

There is, however, a consequence of Popper's analysis that in my opinion has not been taken to heart sufficiently by our community. Criticism and testing are of the essence of our work. This means that science is a fundamentally social activity, which implies that it depends on good communication. In the practice of science we are aware of this, and that is why it is right for our journals to insist on clarity and intelligibility, and why meeting our colleagues at conferences is such an integral part of being a working scientist. But we have hardly begun to take note of this fact in the teaching of science. The priority given to conveying as much material as possible in a limited amount of time means that the teaching of communication skills generally takes a back seat (and often not even that) in undergraduate and postgraduate courses. The consequences of this neglect of an important implication of Popper's teaching are sad: most young (and not so young) scientists find it difficult to talk about their work to the general public and to the media, their lecturing is not always of the best, and even their technical presentations can be hard to follow. So often, when postgraduates write up their PhD theses, these are the first coherent pieces of writing they have done in six or more years. This is hard on them, on their supervisors, and on their examiners. Very gradually the need to teach communication skills is being understood, but it is being acted on only in a very limited way.

This is a brief outline of how one scientist sees Popper's influence today. It leaves out much that is very relevant to our time, such as his inveighing against utopianism in science as much as in politics, his vigour in denouncing illogical and misleading ideas held on authority rather than on evidence, and so on. May we enjoy many more of his characteristic and instructive contributions! □

Sir Hermann Bondi is at Churchill College, Cambridge CB3 0DS, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Karl Popper — an influence shining through.

pulling up one leg and advancing it, and then the other. One leg is labelled 'technology' and the other 'science'." It may be that this vital teaching of his has been implicitly understood by the scientific community since the end of the Second World War, when advances in electronics and the ability to use space platforms provided previously unimagined experimental opportunities. It was Popper who made this dependence explicit and clear.

But perhaps it is his stress on problem-solving as the central activity of humans (indeed, all living organisms can be said to be engaged in problem-solving), which describes so particularly appositely what is done in science. Science is driven forward by unexpected and surprising results emerging from new experiments or by the appearance of contradictions between theories previously thought compatible. Solving such problems as they arise is of the essence of our work. Thus science is not something strange and odd but the most human of pursuits. Hoyle, in his novel

Academic standards in Italy

SIR — G. F. Gaetani and A. M. Ferraris (*Nature* 353, 10; 1991) examined the proceedings of a committee set up by the Italian Ministry of University, Scientific Research and Technology (MURST) to evaluate the scientific achievements of candidates for full professorship in a particular branch of medicine. They were perplexed that the proceedings did not mention the criteria used for evaluation, nor the adoption of a scale of values or of internationally accepted scientific indicators.

The indicators proposed by Gaetani and Ferraris were later criticized by S. Amadori *et al.* (355, 581; 1991) as possibly misleading, while the committee in question volunteered some explanations of the criteria adopted (M. Baccarani *et al.* 356, 188; 1992), whose extravagance, in view of the existing Italian laws, was pointed out by F. Aiuti (356, 556; 1992).

I should like to present evidence that no scale of values for scientific achievement is normally adopted in Italian university competitions for professorships, so that, with the full knowledge and approval of MURST, selection for academic promotion has become increasingly arbitrary and less and less dependent on the scientific accomplishments of the candidates. The evidence is taken from the official proceedings of a recent national competition for tenured positions at the level of associate professor in theoretical and mathematical physics, and thus lends support to the suggestion by G. Visconti (356, 188; 1992) that the present machinery for academic promotion in Italy produces adverse effects well beyond the boundaries of the faculty of medicine, contrary to the belief of F. Aiuti *et al.* (356, 188; 1992) that it produces good results in physics.

In June 1991, a committee of nine theoretical physicists (five full and four associate professors) was set up and several meetings were held during the following few months to reduce the number of applicants to about a third of the original 150.

On 21 November 1991, one member of the committee remarked that even the vague guidelines issued by MURST implied comparison of the scientific achievements of the candidates, in view of the final selection of 14 winners, and proposed the adoption of criteria for making such a comparison. The rest of the committee objected that evaluation of scientific achievement could not be a matter of 'mechanical' procedures; by a majority of eight to one they turned down the proposal and approved a resolution on 16 December 1991 emphasizing the freedom of judgement of the individual committee members.

To clarify the apparent contradiction between the adoption of quantitative indicators and freedom of judgement, the dissenting member then proposed that these indicators be considered together with other criteria (with relative weights to be agreed later) without prejudice to the freedom of judgement of individual committee members. The president of the committee refused to put this to the vote, and officially accused the dissenting member of obstruction. The question, of course, was not about the nature of any particular indicator but whether quantitative criteria should be adopted. It was clearly the opinion of all but one of the committee that preliminary agreement on a scale of values was inappropriate or superfluous.

Between December 1991 and February 1992, the dissenting member wrote three times to the minister, explaining the situation and asking for a clear statement of whether the adoption of quantitative criteria for the evaluation of scientific achievement was permissible or not. The only response from the minister was a telegram of dismissal from the committee on 18 February 1992.

The case illustrates important aspects of the generalized practice in Italy for academic promotion, and also provides a solution to Gaetani and Ferraris's puzzle on the criteria used to evaluate scientific achievements. It also shows that academic authorities at the highest level are fully aware of the prevailing practice.

F. Persico

*Istituto di Fisica
dell'Università degli Studi,
Via Archirafi, 36,
I-90123 Palermo, Italy*

SIR — The proposals of Professor David Burr (*Nature* 357, 273; 1992) for improving the performance of Italian universities make only brief mention of an all-important need, "linking university funding to performance", a point on which one needs to be more explicit. In fact, an increase in the autonomy of the universities, particularly in assigning faculty positions, could be misused by powerful local lobbies in the absence of rigorous evaluation mechanisms for sharing out financial and other public resources.

The absence in Italy of adequate evaluation processes not only for teaching and research in universities but also in many areas of nonacademic science has long been bemoaned both nationally and internationally. In the face of strenuous opposition by interested parties, however, the remedies so far adopted are little more than palliatives. Given these precedents and the present political situation,

the remedies we need, such as the creation of strong advisory bodies with extensive international participation, are unlikely to be forthcoming in the foreseeable future.

In the meantime, problems such as those mentioned previously by Eugenio Tabet and myself (see *Nature* 348, 104; 1990) can only get worse. For example, competition for research funds is becoming more frantic because of an increase in public funding for industrial research and development, at least in certain areas. This political decision and the resulting decrees of the Minister of Universities and Research — for example, the recent one soliciting applications by industry consortia for the assignment of 108 billion lira (about £50 million) for neurobiological research — means that research groups at public institutions must beg for subcontracts from private parties, as a substitute for an evaluation-based financing by less devious pathways.

Giorgio Bignami

*Istituto Superiore di Sanità,
Viale Regina Elena 299,
I-00161 Roma, Italy*

Peer review

SIR — Peer review (*Nature* 357, 354; 1992) has its strengths but it is not perfect. Its weakness is its strength — its susceptibility to feedback.

Peer review is like a black box, the input to which is the attempts of scientists to get published, the output from which is the material that gets published. The black box contains everything necessary for positive and negative feedback.

Neither sort of feedback is here automatically good or automatically bad, it depends on the ends the feedback obtains — and on one's view of what is desirable. The ends run from suppression of an article — or subject — through modification to encouragement. Feedback applied to any given article is, *a priori*, as likely to benefit the reviewer as it is the author, as likely to benefit the reviewer as it is the readership, as likely to benefit the reviewer as it is mankind.

To say that something possesses a capacity is not, however, to say that that thing makes continual and untrammelled use of that capacity. It does not automatically follow that either sort of feedback current flows in any one instance of the peer review circuit. The resistance preventing current flowing — or preventing an undesirable sort or amount of current flowing — seems simply to be constituted by the judgement of the individuals concerned.

J. A. Negus

*Bessas,
07150 Vallon Pont d'Arc, France*

Counting the numbers

SIR — I was shocked to read that *Nature* is to devalue the billion in line with American usage (358, 2; 1992). After all, not everybody who reads a copy of an article will have read this article too. He also has to remember that you made up your mind on 2 July 1992. Instead of increasing the confusion with this decision, it would perhaps be better to find an unambiguous way of citing such figures: The decimal prefixes are excellent tools for this purpose. If you use M = mega for 10^6 , G = giga for 10^9 , and T = tera for 10^{12} you find that \$5G is quite distinct from \$5T. You could even use P = peta for 10^{15} or E = exa for 10^{18} knowing that you are acting according to the international standard ISO 1000 and that there is no room for misunderstanding. These decimal prefixes are also useful in tables when the range for figures is sometimes very extended.

Martin A. Lobeck

Auf der Stippe 19,

D-5222 Morsbach-Hülstart, Germany

SIR — So *Nature* has abandoned the English use of billion (10^{12}) in favour of the American use (10^9). I find the excuse to be rather feeble: "... it has been difficult to dragoon correspondents into describing, say, the US federal deficit as 'close on \$500,000 million' ..." Who ever tried to dragoon them into that was mistaken: the correspondents should have been dragoonned into describing it

as 'close on \$500 milliard'.

P. J. A. Howard

*Ashrigg, Ashmount Road,
Grange-over-Sands,
Cumbria LA11 6BX, UK*

SIR — I was born in Europe and lived many years in America. As American mathematicians and scientists use the terms million 10^6 and billion 10^9 , I (a drop in the ocean) am prepared to accept it. I was taught at school that million is 10^6 , billion 10^{12} , trillion 10^{18} , and the numbers in between are called milliard 10^9 ,illiard 10^{15} .

L. G. Manittus

*Sct. Pedersgade 2A,
3700 Rønna,
Bornholm, Denmark*

SIR — Why should not a scientific journal use the SI prefixes? Megabuck is already in colloquial use and a new precision might be indicated by the use of megadollar and so on. The only disadvantage is that the teradollar might be heard to refer to 'dollar imperialism'.

D. P. Hansell

*30 The Ridgeway,
Tonbridge, Kent TN10 4NH, UK*

□ The usage (which *Nature* has now abandoned) that 1 billion = 10^{12} was part of the formal recommendation of the Conférence Générale des Poids et Mesures in 1948. — Editor, *Nature*.

Building a low-cost space station

SIR — Last year, many in the US science community united in opposition to space station Freedom and argued that its enormous cost of \$30 billion would not produce an adequate scientific return. After an extended and highly charged debate, Congress agreed to continue its funding. This year, budgetary difficulties are mounting and Congress is engaged in another debate. Not included in the discussion so far is an alternative space station design proposed to the National Aeronautics and Space Administration last year, which would cost less and entail fewer risks than the existing shuttle system¹.

The proposed space station, which we call STS-Lab, consists of two main elements from the existing shuttle: a modified orbiter (Orbiter-II) and a modified external tank (ET). All the re-entry and landing flight systems are removed from Orbiter-II including the wings, tail, thermal protection tiles, landing gear and avionics, reducing the 180,000-pound weight by 40,000 pounds.

The Orbiter-II cargo bay contains a

laboratory module and a solar-powered extended-duration orbiter (SPEDO) power system² providing 18 kW of average power. Spacelab or one of the modules under construction for space station Freedom could be used for the laboratory module. The ET intertank volume has pressurized access tunnels connecting the ET pressure vessels, the Orbiter-II mid-deck and several docking ports. A visiting orbiter, the European and Japanese laboratory modules being built for Freedom and a crew rescue vehicle could be accommodated by attachment to these docking ports.

We estimate that STS-Lab can be built for \$3 billion within four years. It would be placed in orbit in a single unmanned launch; our cost estimates assume the conversion of Columbia, modification of an existing ET and use of an existing Freedom laboratory module. The STS-Lab design and the cost and schedule estimates were developed by a group of experienced aerospace engineers.

In general, there are no significant differences in the research that could be

supported by Freedom compared with STS-Lab. The same types and amount of scientific experiments could be conducted in either of the orbiting laboratories. However, STS-Lab's large ET volume would allow laboratory space to be easily expanded.

Budgetary support for all categories of science is being threatened by Freedom. As scientists we must do all we can to ensure that the debate in Congress includes serious discussion of practical alternatives.

Randolph H. Ware

*University Navstar Consortium,
PO Box 3000,
Boulder, Colorado 80307, USA*

1. Ware, R. H. & Culbertson, P. E. *J. pract. Appl. Space* **3**, 17–23 (1992).

2. Garriott, O. K. *J. pract. Appl. Space* **3**, 25–32 (1992).

How rare is rare?

SIR — It is commonly said that fraud in science is rare. What is truly rare is to find some objective statement of what the word "rare" means in the context of fraud. In a letter to *ASM News* (August 1991, p. 391) I reported on a survey of 24 scientific colleagues which suggested that to them "rare" means that "the instances of deliberate fraud in basic research" would be no greater than 1 in 10,000 publications. Now, in an article (*Nature* **357**, 427; 1992) on a fabricated paper from the National Institutes of Health, you quote Michael Green of the University of Massachusetts as suggesting that the instances of fraudulent manuscripts may actually be as low as 1 in 100,000. I would repeat in ironic form the challenge I made in my previous letter: can anyone making claims published in a scientific journal about the rarity of fraud in science show us their data?

Irwin Tessman

*Department of Biological Sciences,
Purdue University,
West Lafayette,
Indiana 47907, USA*

Take 100 lines

SIR — Daedalus¹ does not seem to have done his homework. The technique he describes is the well known 'Lempel-Ziv-Welch' data compression algorithm^{2,3}. Perhaps this is an example of morphic resonance?

A. P. R. Cooper

*British Antarctic Survey,
High Cross, Madingley Road,
Cambridge CB3 0ET, UK*

1. *Nature* **358**, 22 (1992).

2. Nelson, M. *Dr Dobbs J.* **156**, 29–36 (1989).

3. Ziv, J. & Lempel, A. *I.E.E.E. Transactions on Information Theory* (May 1977).

Problems for foreigners in Japan

SIR — Japan is attracting more foreign researchers, although the total number of foreigners is still dwarfed by their Japanese counterparts studying or working overseas (see *Nature* 350, 4; 1991). Monbusho (Ministry of Education) opened the doors of national universities to foreign lecturers in the 1970s after a special law was passed by the Diet, and a more significant law in the early 1980s allowed national universities to employ foreign professors and assistant professors on a permanent basis. In 1986, another progressive amendment allowed foreigners to be employed by the 84 national laboratories. And an increasing number of young foreign researchers are applying for the 200 or so postdoctoral positions in national institutes, in addition to fellowships available within the university system.

Despite these impressive relative gains by ministries with research agencies, Homusho (the Ministry of Justice) is actually moving in the opposite direction, with a more restrictive immigration policy. In July 1990, the law governing the granting of visas was modified, with foreign workers in Japan (in all fields, including research) being eligible for a maximum one-year work permit, in contrast to previous regulations which allowed up to a three-year work permit. Exceptions include medical doctors and lawyers, as well as employees of Monbusho (university lecturers), who may receive up to three-year work permits. At the same time, there is no consistent policy in granting permanent residency. Regardless of apparent regulations, applications are accepted only after a lengthy period on a work permit.

Although a Japanese national institute may offer a foreign researcher a three-year to lifetime contract, this security is eroded when Homusho offers only single-year visas. This negates all the other efforts being made to integrate foreign workers into the Japanese system, and serves only to encourage foreign researchers to stay home or seek employment elsewhere.

My situation illustrates the inconsistencies a foreigner is faced with. In 1982 I received permanent residency in New Zealand when I joined a Department of Science and Industrial Research (DSIR) national laboratory. This was standard practice, with more than 10 per cent of my DSIR colleagues having been born in Asia, Europe, North America and elsewhere. In April 1989, I moved to Japan to take up a three-year appointment in a national institute of the Agency of Industrial Science and Technology (AIST) of the Ministry of International Trade and Industry (MITI). In my first three

years, I was very satisfied with the progress of my research, as well as my working conditions, colleagues and life in Japan. When my contract came up for renewal I was keen to continue, but I requested that there be no expiration date on my new *jirei* (letter of appointment), the same as that received by all my Japanese colleagues. This was granted and I became the first foreigner in Japan's national institutes to hold a *jirei* without an expiration date, thus truly joining the permanent staff.

At the expiration of my initial three-year visa, I requested from the Immigration Bureau another three-year permit and also inquired about permanent residency. Despite my permanent status with a national institute, I was told I could receive only a one-year visa and that I could not yet apply for permanent residency; I might have to wait perhaps a further nine years. The permanent residency aside, I pointed out that Monbusho employees qualify for a three-year visa, often with contracts for specific periods in contrast to my permanent position. The officer agreed that this seemed inequitable, but nevertheless was the law.

In a recent interview by Akwi Seo with Immigration Bureau staff of Homusho (*Look Japan* 37, 32; December 1991), this apparent inequality between staff of national universities and national institutes was explained: "... professorships are difficult to obtain, entail a high degree of social status, and aren't accessible to 'just anybody'", implying that both Japanese and foreign national institute staff are of lower status. Seo also interviewed a civil-rights lawyer for foreign residents, who said that the agencies that hired foreigners "have failed to exert pressure on the Ministry of Justice, leaving their foreign employees at the mercy of the immigration services", which "prefer issuing short-term visas because they require frequent visits to the Immigration Bureau, helping the Ministry [of Justice] keep track of Japan's growing foreign population". Although some agencies may be slow to advocate the position of their foreign employees to Homusho, AIST has been supportive in my case, but to no avail. The present situation simply reflects the highly conservative nature of Homusho, and the fact that a ministry can and often does ignore requests originating from other ministries.

The case of foreign researchers is only a small part of the much wider issue of foreign labour in Japan. With relatively little experience in dealing with foreigners living and working side by side with Japanese, Japan may benefit from

outside views in dealing fairly and consistently with its growing foreign population. Bureaucrats of such a homogeneous society must realize that they can never fully appreciate what it is like to be a *gaijin* (outsider) in Japan, and the problems we face daily in trying to live and work here, a country many of us choose to call our home. We simply ask for equitable consideration, in a world perspective.

Jeffrey W. Hedenquist

*Geological Survey of Japan,
Tsukuba, Japan*

Nineties' research

SIR — At a recent meeting of the Departmental Nomenclature Committee (DNC), it was agreed to adopt the following minimum set of standards:

(1) All academic members of staff are to be named professor. Individuals should think of suitable titles for themselves, and colleagues should be discouraged from the habit of addressing each other by first names. All other staff should be called research associates, with subtle variations (research associate-in-chief, assistant associate).

(2) All rooms are to be renamed research units: the instrument room is to be called the photonic detection research unit, the dark room the luminosity minimized quantal imaging unit and so on. It is essential that visitors (potential patrons, donors and the rest) should see only the label on the door.

(3) Research groups are to be renamed molecular research groups. The term genetic should be dropped from group titles as market research shows that this word conjures up pictures of bespectacled old men pottering among pea plants obtaining questionable results, a corporate image hardly likely to attract external funds.

These changes should bring the department into line with the other members of the European Communities (making applications for grants more likely to be successful), should impress the public at large, boost our morale and help us to develop a credible corporate profile.

Simon Wolff

*Professor of Molecular
Biohazard Evaluation,
Molecular Environmental
Bioepidemiology Research Unit,
Suite 3.20-3.21,
Third Plaza (Mezzanine),
Division of Molecular Pharmacology
and Molecular Toxicology,
Department of Molecular Medicine,
Molecular Life Sciences Consortium,
University College London,
5 University Street,
London WC1E 6JJ, UK*

Humbling of world's AIDS researchers

Last week's international conference on AIDS seems merely to have confirmed the impression of earlier conferences in the series — that science, for the time being, has little to say to those with AIDS.

LAST week's conference in Amsterdam seems to have had more in common with the meeting there in May (see *Nature* 357, 188; 1992) than the organizers planned, or bargained for. In particular, as newspaper accounts of what happened have diligently recorded, some sessions were marked by open confrontation between people diagnosed as suffering from AIDS and researchers doing everything they can to understand this distressing condition and, eventually, to find ways of treating it. Drug companies such as the Wellcome Foundation (Burroughs-Wellcome in the United States) seem to have been especially ill-treated for making money out of AZT, the drug most often used for restraining the progress of the disease.

None of that is surprising. People in whom immune deficiency has already appeared, even if recognized only by a T-cell count, may feel that they necessarily exist on the margins of ordinary living space, never sure when they will be struck down by some adventitious infection. It is understandable that they should be enraged by what may seem a cruel fate, made no easier to bear by the inclination of others to regard them as outcasts. That many of those in this condition may also be united with groups of like-minded people by practices that may have contributed to their contracting the disease in the first place makes it inevitable that individual rage will be collectivized. It is not reasonable to expect them to behave as saintly Christian martyrs.

That much of the rage may be directed against researchers is also unsurprising. When AIDS was first distinguished as a disease entity more than a decade ago, consternation was appropriate. By 1984, when the association of AIDS with HIV (as the virus is now called) had been recognized, it was forgivable (if injudicious) of Mrs Margaret Heckler, then the US Secretary of Health and Human Services, to speculate about the production of a vaccine.

That promise, for one, has not yet materialized. Whether it ever will remains to be seen. It is good news, from last week, that several vaccines have already been tested on volunteers not infected with HIV, without untoward consequences. It will be some time longer before anybody can tell whether such agents are likely to be useful in the treatment of those already suffering from AIDS, or whether they can be used only prophylactically, and presumably only in the rich countries of the world.

Other promises widely understood to be implicit in the biomedical research enter-

prise have similarly failed to materialize. Why is there no drug (except AZT, and its few analogues now making their way into clinical practice)? Why is there no magic bullet, perhaps a ribozyme, specific in its action against HIV, which leaves other nucleotide sequences intact? And if there is, as some say, an intimate connection between the efficacy of the immune system and the working of the pituitary gland, as proxy for the mind's ambitions, why have people not yet devised the way of thinking one's way out of trouble caused by AIDS?

Sadly, there is no one-to-one relationship between disease entities and drugs that will cure them. The hunt for magic ribozymes has been tried unsuccessfully, and will no doubt be tried again (but drug delivery will not be easy). And psychoimmunology, despite the benefits there would be if it could be demonstrated to be other than an autonomous connection, has been ineffectual in dealing with far simpler diseases than AIDS.

The decade's gap between the recognition of AIDS and the present is not in itself remarkable, of course. Tuberculosis may have been recognized to be an infection soon after Pasteur, when measures such as the relative isolation of patients and the exercise of their lungs helped reduce the death toll, but effective measures against the disease *in situ* have been available only for the past half-century.

Most infections more recently recognized as such have been more quickly dealt with. The successful containment of poliomyelitis, even in the poor countries, is only the most dramatic of a string of successes. By 1984, it had become everybody's instinct to suppose that most infections were curable one way or another. One of the reasons why the research community is now pilloried by those with AIDS is that much the same was promised, or was assumed, for the treatment and even cure of AIDS.

But why blame the research community for that? The question, however plaintive, is reasonable enough. The trouble is that it cannot confidently be answered in the sense researchers would wish. While everybody accepts that there are intellectual problems about the world we live in that are for the time being insoluble, there is a tendency to suppose that mere practical problems — fixing a machine or curing a disease — can be simply dealt with. One lesson usefully to be learned from the the past decade's disappointment over AIDS is that researchers should be more diligent in the repetition of what everybody knows — that uncertainty

is always with us. The collective anger of the activists at Amsterdam may be a lasting reminder of that.

Last week's conference seems to have been humbling in another way; this year, there was no great leap forward in understanding to report. No doubt that is why so much attention was paid, at least by European daily newspapers, to reports of people apparently suffering from AIDS in whom no trace of HIV can be found. Does that mean that Duesberg has been right all along, and that HIV plays no part in the causation of AIDS?

Mercifully, most people appear to have appreciated that that conclusion would be false. For one thing, it is too soon to know what will be made of these observations when eventually they are published, some apparently in the next week or so. But there are several explanations, one of which is that severe (and, if untreated, fatal) immunodeficiency long predates the discovery of HIV, but is evidently distinct from the rash of cases that made their first appearance in San Francisco in the late 1970s. Awareness of AIDS (and the ease with which T-cell status can now be assessed) could easily have helped to bring cases of that kind to the attention of physicians. But it is also possible that cases of AIDS without traces of HIV reflect circumstances in which the virus has done its nasty work by triggering a secondary failure of the immune system and has then itself been eliminated from the cells at risk. And, of course, there is always the possibility of experimental error.

What will happen next? Many of the participants in last week's meeting found the occasion enervating, while the views of those who still believe that something dramatic will turn up sound more and more Micawber-like. On balance, the view that there is a long haul ahead predominates.

Participants are also divided on the question whether, at future conferences in this series, there should be an attempt to separate the presentation of research work from the educational function of enabling AIDS organizations to mix with those in whose hands their clients' lives may rest. While some researchers would prefer a setting in which they could understand what their peers are saying, more would prefer not to be thought to have ducked confrontation.

Next year's meeting has been arranged for Berlin, after which the conferences will probably happen every other year. How long will it be before there is a hopeful conference?

John Maddox

The germ of the issue

Susan Strome

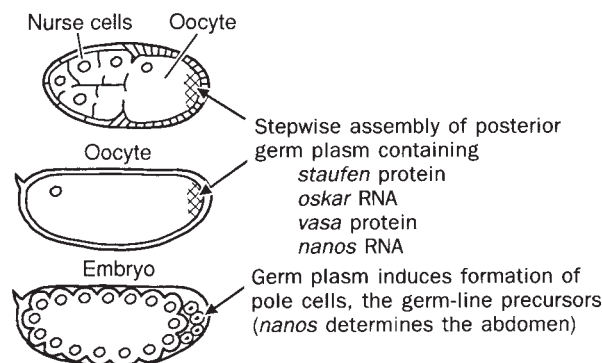
Drosophila is in the vanguard of the attack on the problems of cell-fate specification and pattern formation, and the article by Ephrussi and Lehmann on page 387 of this issue¹ represents another notable advance. The authors demonstrate that the level of the *oskar* gene product determines the number of germ cells formed in *Drosophila* embryos, and that mislocalization of *oskar* RNA is sufficient to induce ectopic germ-cell formation. The study establishes the importance of *oskar* and clarifies the roles of several other maternal products in specification of germ-cell fate.

As in many organisms, the germ line is set aside early in *Drosophila* embryogenesis (see figure). Ever since the determinative properties of the posterior 'germ plasm' were demonstrated by cytoplasm transplantation experiments², the quest for a 'germ-line determinant' has been intense. As with many other searches for key developmental molecules, the genetic approach is proving most fruitful. In *Drosophila*, genes encoding components required for germ-line development have been identified in screens for maternal-effect lethal mutations that alter pattern formation during early embryogenesis³. Mutations in eight genes (*cappuccino*, *spire*, *staufer*, *oskar*, *vasa*, *tudor*, *valois* and *mago nashi*) result in failure to form the germ-line progenitor cells and in abnormal abdomen development^{3,4}. Mutations in *nanos* and *pumilio* also cause defective abdomen development but do not affect the germ line^{3,4}.

Pathway

How do these gene products involved in posterior pattern formation interact, and what part do they serve in germ-line and abdomen development? The picture emerging from genetic and molecular studies^{1,5-9} is that germ-line and abdomen development are controlled by two distinct signals: *nanos*, the abdomen signal⁵, and an unidentified germ-cell signal. Both signals must be localized to the posterior end of the early embryo. The eight genes required for both germ-line and abdomen development seem to be involved in transport and localization of the two signals and assembly of functional germ or pole plasm^{1,6-9}. The known steps in this localization/assembly pathway are these: *cappuccino* and *spire* are required for posterior localization of Staufen protein; *cappuccino*, *spire* and

staufer are required for posterior localization of *oskar* RNA; and *cappuccino*, *spire*, *staufer* and *oskar* are required for posterior localization of Vasa protein. The *tudor* and *valois* genes are not required for any of these events, tentatively placing them downstream in the pathway. So far, analysis of this hierarchy of assembly has not revealed whether any of the gene products is the germ-line signal. The study of Ephrussi and Lehmann does, however, eliminate



Assembly of germ-line-inducing pole plasm during *Drosophila* oogenesis, and induction of pole-cell formation at the syncytial blastoderm stage of embryogenesis.

several genes from contention.

To address which gene products are sufficient for germ-line induction, Ephrussi and Lehmann caused mislocalization of *oskar* RNA to the anterior end of embryos by transforming flies with the *oskar* coding sequence fused to the anterior localization signal in the 3' untranslated region of the *bicoid* gene. Females containing the fusion gene produce embryos with two striking defects: the head and thorax are replaced by a mirror-image duplication of the posterior abdomen (because of ectopic *nanos* localization at the anterior pole), and germ cells are formed at both the anterior and posterior poles of the embryo. Tests of this 'anterior *oskar*' effect in various mutant backgrounds show that *cappuccino*, *spire*, *staufer*, *valois* and *mago nashi* are not required for anterior germ-cell or abdomen formation, whereas *tudor* and *vasa* are.

Except for the result with *valois*, these findings are consistent with the assembly pathway described above. Furthermore, they eliminate *cappuccino*, *spire*, *staufer*, *valois* and *mago nashi* from consideration as germ-line signals. Evidently, functional pole plasm, containing the germ-line and abdomen signals, can assemble anteriorly in the absence of several of the gene products that are required for posterior pole-plasm

assembly. Because this relies on anterior anchorage of *oskar* RNA, the likely functions of the dispensable gene products are localization and anchoring of components of the pole plasm to their usual site at the posterior pole.

These results focus our attention on *oskar*, *vasa* and *tudor* as essential for germ-line induction. Further tantalizing evidence that *oskar* has a special part in the process comes from the observation that increasing the dose of the normal *oskar* gene leads to an increase in posteriorly localized *oskar* product and to an increase in the number of primordial germ cells¹. Increasing *oskar* dose also leads to elevation of posteriorly localized *vasa*, *staufer* and *nanos* product. This implies that the product of *oskar* is a limiting factor in germ-plasm assembly. Whether it is an essential factor or one whose function can be bypassed can be addressed by testing the ability of mislocalized *vasa* or *tudor* product to induce ectopic germ cells in different mutant backgrounds.

Germ granules

One of the questions that remains is whether ectopic germ plasm contains the distinctive granules normally seen in the posterior germ plasm. The presence of these electron-dense granules in the germ-line cytoplasm of organisms as diverse as insects, nematodes and amphibians¹⁰ has prompted many investigators to speculate (and several textbook authors to state) that the granules contain the crucial germ-line signal. Two observations bolster this hypothesis: mutations in the eight genes that affect both germ-line and abdomen development disrupt germ granules^{3,4}; and Vasa protein is a component of the granules⁹. Determining whether ectopic germ plasm contains germ granules, and whether embryos with increased levels of *oskar* product contain an increased number of germ granules, may shed light on the germ-granule issue.

The pathway for germ-line and abdomen induction has been whittled down to *oskar* → *vasa* → *tudor* → germ cells, and *oskar* → *vasa* → *tudor* → *nanos* → abdomen (see Fig. 6 of Ephrussi and Lehmann's paper, page 391). A reasonable model is that the products of *oskar*, *vasa* and *tudor* provide a scaffold, perhaps in the form of germ granules, for the abdomen signal (*nanos*) and a germ-line signal. The dual function of a germ-line scaffold to anchor both signals at the posterior pole would explain why all of the maternal-effect, germ-line-required genes identified in *Drosophila* also affect the abdomen and are therefore maternal-effect lethal.

One approach to identifying germ-line specific genes downstream of *tudor* is by screening for maternal-effect sterile mutations. Although such screens are difficult in *Drosophila*, they can be done¹¹ and could identify the germ-line signal gene or else verify that a gene that has already been identified encodes the signal. Screens for maternal-effect sterile mutations in *Caenorhabditis elegans* have identified five genes required for germ-line development¹². Most pertinent to the issues discussed here, the *C. elegans* mutations do not disrupt germ granules and seem to affect the germ line only. Whether the differences between maternal, germ-line-required genes in flies and worms reflect differences in the types of mutations screened for, different functions of germ granules in the two species or different mechanisms of germ-line specification, are matters that remain to be determined.

Insights into germ-line development are coming not only from *Drosophila* and *C. elegans*, but also from *Xenopus*. When primordial germ cells are transferred to ectopic locations, they do not develop into germ cells but instead join other tissues and differentiate into a variety of somatic cell types¹³. This shows that *Xenopus* germ plasm, with its associated germ granules, does not irreversibly commit the cells that receive it to a germ-line fate and that the correct environment is crucial for germ-cell differentiation. Indeed, all studies in *Drosophila* in which ectopically formed germ-line precursor cells have been proven to be functional have involved transferring them to a normal location for migration to the somatic gonad^{1,2}.

So, whatever the germ-line signal turns out to be, it may not be a 'determinant' in the strictest sense of the word — the determination of germ-cell fate may require other environmental cues or may be easily over-ridden. Indeed, the germ signal may not instruct primordial germ cells so much as protect them from somatic differentiation signals¹³. □

Susan Strome is in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

An air-conditioned greenhouse

Graeme Stephens and Tony Slingo

WITH the oceans assuming an ever greater significance in our understanding of climate, there is obvious interest in the claim of Ramanathan and Collins¹ that there is a natural thermostat which prevents sea surface temperature (SST) rising above 305 K. The mechanism, which relies on extensive cirrus cloud cover put up over the warmest waters to reflect solar radiation, is challenged by Fu *et al.* on page 394 of this issue². They argue from satellite data that changes in cirrus cover are related more to changes in atmospheric circulation than to SSTs, and that changes in surface evaporation are more important than Ramanathan and Collins allow. This evaporation provides continuous air conditioning of the tropical SSTs, reducing the need for any thermostat.

That Ramanathan and Collins's study should be provocative is not surprising, as the topic of climate feedbacks is highly controversial. Although it is attractive to use observations to study feedbacks, it is often impossible to separate cause and effect. With innovative use of observations obtained before and during the 1987 El Niño climate fluctuation in the Pacific, Ramanathan and Collins tried both to isolate and to quantify the cloud feedback. They propose a simple relationship between changes in sea surface temperatures, cirrus clouds and solar radiation, and assume that other processes are of secondary importance. But can cause and effect be established so clearly in the equatorial Pacific, where the conventional wisdom is that SSTs are maintained by many coupled air-sea interactions and ocean-atmosphere transport processes, operating on a variety of time and space scales^{3,4}?

Several processes are known to play important roles in determining the SST distribution. In his critique of the idea, Wallace⁵ articulated the common belief that evaporation increases with SST and so acts to smooth out local SST maxima. But recent studies⁶ show that evaporation is not necessarily higher over warmer SSTs. Ramanathan and Collins note⁷ that the climatological evaporation decreases along the equator from the east Pacific (regions of cooler SSTs) to the warm pool region where the SSTs are a maximum, contrary to what may be expected from Wallace's arguments. But this observation does not diminish the importance of evaporation, either to the heat budget of the equatorial Pacific as a whole, or to the SST anomalies that occur during El Niño. This is supported by studies of the role of transients such

as the intra-seasonal oscillation and cold surges⁴ and by recent model simulations of El Niño⁸.

Large-scale motions also play a significant role in governing the observed distributions of cloudiness and SST. The evolution of El Niño is thought to be influenced principally by shifts in the circulation patterns of both the atmosphere and oceans^{3,4,8}. This strong dynamical influence is perhaps the most fundamental difference between Ramanathan and Collins and Fu *et al.*, who support Wallace's belief that large-scale dynamical processes spread local influences to other regions of the tropics and that analyses like that of Ramanathan and Collins need to consider this larger domain. They show that the correlation between SST and cloudiness breaks down on the largest scales, which they believe are the most important. Ramanathan and Collins argue that their thermostat operates locally, over the highest SSTs, with atmospheric motions invoked to export the longwave cloud forcing from the region. This must require a selection process, as this feedback does not apply to all regions of high SST. For example, over the western Pacific during El Niño the SSTs are still near their maximum values, but there may be no deep convection or cloudiness.

Precipitation is another important process relevant to tropical SSTs. Precipitation exceeds evaporation in the west Pacific, providing an influx of fresh water into the ocean. This reduces the salinity of the upper few metres of the mixed layer, creating a stable, less-dense layer of water at the surface (the so-called fresh-water lens). This stable layer resists wind driven mixing with the colder underlying water and is heated by the absorption of solar radiation⁹, thus further stabilizing the layer. Such a warmed layer promotes further convection, more rainfall and further warming. Perhaps Ramanathan and Collins's thermostat operates only over such regions of significant fresh water input. Alternative mechanisms^{3,4} to constrain SSTs in these regions are the transients mentioned above and the episodic mixing events driven by westerly wind bursts that are strong enough to access the colder water below the fresh lens, thereby cooling this layer and limiting the SSTs.

The data from the satellite-based Earth Radiation Budget Experiment (ERBE), used by Ramanathan and Collins, show a remarkable and as yet unexplained near-cancellation over tropical convection between the large cooling

1. Ephrussi, A. & Lehmann, R. *Nature* **358**, 387–392 (1992).
2. Ilmensee, K. & Mahowald, A. P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1016–1020 (1974).
3. Nusslein-Volhard, C. *et al.* *Science* **238**, 1675–1681 (1987).
4. St Johnston, D. *et al.* *Cell* **68**, 201–219 (1992).
5. Wang, C. & Lehmann, R. *Cell* **66**, 637–647 (1991).
6. Kim-Ha, J. *et al.* *Cell* **66**, 23–35 (1991).
7. Ephrussi, A. *et al.* *Cell* **66**, 37–50 (1991).
8. St Johnston, D. *et al.* *Cell* **66**, 51–63 (1991).
9. Hay, B. *et al.* *Development* **103**, 625–640 (1988).
10. Eddy, E. M. *Int. Rev. Cytol.* **43**, 229–280 (1975).
11. Boswell, R. E. *et al.* *Development* **113**, 373–384 (1991).
12. Capowski, E. E. *et al.* *Genetics* **129**, 1061–1072 (1991).
13. Wylie, C. C. *et al.* *Dev. Biol.* **112**, 66–72 (1985).

due to the reflection of solar radiation back to space and the equally large greenhouse warming due to trapping of upwelling thermal radiation. As seen from space, the net effect of clouds is therefore close to zero, but the heating has been redistributed significantly within the column¹⁰. Ramanathan and Collins recognize that the cooling effect is felt at the surface, whereas the greenhouse warming occurs within the atmosphere. They claim that this warming is advected away by the atmospheric circulation and does not influence the local SSTs, which experience only the cooling effect. However, modelling studies^{10,11} show that the atmospheric warming induces changes in the vertical motion field and thereby influences both the precipitation and circulation fields, which in turn influences the cloud cover at all levels.

Our understanding of the processes that maintain the warmest SSTs on the planet is rudimentary. Compared with the tropical eastern Pacific, the western Pacific is a region where the winds are generally light and the evaporative heat flux is small, where the structure of the ocean mixed layer is complex and the SST is highest and where large-scale east-west and meridional circulation of the atmosphere ensures significant moisture convergence to support the deepest convection, heaviest rainfall and smallest fluxes of solar radiation into the ocean. How these processes intimately link together is not well understood. We have attempted to expose only the simplest aspects of these coupled processes and suggest that the longwave cloud forcing, the precipitation and evaporation as well as dynamical influences on cloud cover must also be assessed to determine whether Ramanathan and Collins's thermostat can function. It is our view that it is impossible to isolate a given feedback unambiguously from other coupled processes using available observations.

Nevertheless, the thermostat and the air-conditioning mechanisms are tantalizing hypotheses which perhaps can only

be tested fully by models of the coupled ocean-atmosphere system, once these models have demonstrated sufficient realism. To this end, improved basic understanding of the processes of evaporation, precipitation and radiative transfer and how they couple, together with a better understanding of the influence of large-scale atmospheric and oceanic circulation on these processes, is needed urgently. So it is most timely that a major international experiment will be conducted later this year in the west Pacific; the Tropical Oceans and Global Atmosphere Coupled Ocean-

Atmosphere Response Experiment (TOGA-COARE¹²). This will provide some of the sophisticated new observations which are needed to improve our understanding of the processes which maintain the warm sea surface of the equatorial Pacific. □

Graeme Stephens is in the Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado 80523, USA. Tony Slingo is at the Hadley Centre for Climate Prediction and Research, Meteorological Office, Bracknell RG12 2SY, UK.

ORE GEOLOGY

A little rain must fall

Andrew P. Gize

DEEP below the permafrost of northern Alaska, rainwater (meteoric water) which fell perhaps as long ago as 10^5 – 10^6 years over the Brooks Range mountains in the centre of the State, is carrying heat and minerals through aquifers to the northern coastal plains. In showing this, using detailed borehole studies, Deming *et al.*¹ have confirmed the geochemically and economically important roles that long-range groundwater migrations can have in large sedimentary basins. In addition, their data provide a solid framework against which many geological and geochemical concepts can be evaluated.

The significance of groundwater flow in sedimentary basins is that it provides a fundamental means of transferring heat and mass. The consequences are varied, and include control over the extent of mineral equilibrium or disequilibrium assemblages in the sediments, the force responsible for petroleum secondary migration, and control over the transport and location of some types of metalliferous ore deposits.

Several interpretations have been put forward to describe the sources and driving forces for the water. Some water could be expelled during clay mineral structural transformations or by compaction from overlying sediments². Alternatively, water could be driven by meteoric water precipitated on topographically high regions, providing a hydraulic head³, or by density gradients resulting from thermal expansion or salinity gradients which then induce free convection⁴. Each of these alternatives has merit in certain circumstances. Neverthe-

less, a generalized observed pattern has emerged for groundwater flow in basins bordering mountain ranges¹. In the mountain-range foothills, new meteoric water recharge depresses the geothermal gradient and surface heat flow as it moves downwards. In the basin centre, groundwater flow tends to be horizontal, with little variation in the vertical heat flow. At the far end of the basin, the fluid flow, now rising, is controlled by details in the basin geometry, and localized highs arise in the geothermal gradients and surface heat flow.

The results of Deming *et al.* are the culmination of a 10-year collaborative programme of the US Geological Survey and the National Petroleum Reserve Alaska, covering an area of approximately $300 \times 500 \text{ km}^2$ between the Brooks Range mountains and Prudhoe Bay on the northern coast. The borehole data were obtained between 1977 and 1984, and were used to confirm the generalized heat flow model for the basin⁵: in the Brooks Range foothills the vertical geothermal gradients are less than 22°C km^{-1} whereas on the coastal plain they are as high as 53°C km^{-1} . Although it was suggested then that the

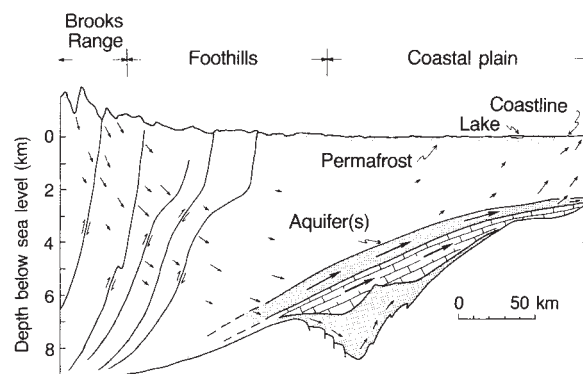


FIG. 1 Conceptual model of regional groundwater flow in the North Slope of Alaska (from ref. 1).

1. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).
2. Fu, R., Del Genio, A. D., Rossow, W. B. & Liu, W. T. *Nature* **358**, 394–397 (1992).
3. Webster, P. J. in *Tropical Extra-Tropical Interactions*, 67–116 (ECMWF, 1991).
4. Murakami, T. *Meteorol. atmos. Phys.* **44**, 85–100 (1990).
5. Wallace, J. M. *Nature* **357**, 230–231 (1992).
6. Sui, C.-H., Lau, K.-M. & Betts, A. K. *J. geophys. Res.* **96**, 3151–3163 (1991).
7. Ramanathan, V. & Collins, W. *Nature* **357**, 649 (1992).
8. Barnett, T. P. *et al. J. Clim.* **4**, 487–515 (1991).
9. Lukas, R. in *International TOGA Scientific Conference Proceedings* 73–81, WCRP-43 (World Meteorological Organisation, 1990).
10. Slingo, A. & Slingo, J. M. *Q. J. R. Meteorol. Soc.* **114**, 1027–1062 (1988).
11. Randall, D. A. *et al. J. atmos. Sci.* **46**, 1943–1970 (1989).
12. US TOGA-COARE Science Working Group *TOGA-COARE Science Plan* (Uni. Corp. for Atmos. Res., Boulder, 1989).

measured geothermal gradient variations could be due to vertical advection, effects from variable sedimentary thermal conductivities were acknowledged to be a potential factor. Now, with thermal conductivity measurements completed, and potential effects from variations in sediment conductivity clearly eliminated, Deming *et al.* show that for the North Slope of Alaska, the only viable interpretation for groundwater movement is forced convection by a topographically driven system (Fig. 1). It is this clear elimination of all other possible mechanisms for the fundamental circulation which allows constraints to be imposed on a wide spectrum of other sedimentary-basin processes.

The significance of basinal groundwater flow in providing a driving force behind the secondary migration of petroleum is discussed by the authors, with the suggestion that the North Slope groundwater flow was involved in petroleum accumulation in Prudhoe Bay. There remain other critical problems in sedimentary basins, several of which can be exemplified by the economically important Mississippi Valley type (MVT) lead-zinc sulphide deposits. In spite of appearing, from their mineralogy, geology and geochemistry, to be among the simplest types of metalliferous deposits, aspects of their genesis from such large-scale basinal groundwater flows remain a source of contention⁶.

A consequence of topographic recharge is that the water input will fluctuate with climatic cycles. The question is how far into the basin these fluctuations are observable before being dampened out. An indication of the fluid flow complexities which are evident at the far end of the basin, and which need to be incorporated into any basin flow model, are indicated by the delicate micrometre-scale banding observed in some MVT sphalerites (such as the Upper Mississippi Valley District, Fig. 2). This banding can be correlated between individual ore bodies on the kilometre scale, indicating large-scale spatial and temporal fluid fluctuations⁷.

Using time-series analyses, some but not all of the banding has been shown to occur at periodicities comparable to known climatic cycles (such as the Milankovitch cycles), and has been attributed to reduced basinal fluids interacting with cyclic, oxidized, near-surface waters⁸. Similar banding in petroleum reservoir silicates cannot be related to interaction with shallower waters, and reflects either fluctuations in the recharge area or structural processes in the basin^{9,10}. Clearly the interpretation presented by Deming *et al.* will be far more complex in reality with fluid flow fluctuations arising at several points. One intriguing question is how far along

the flow path can these fluctuations be observed, and what are their relative contributions at the end of basin flow?

Deming *et al.* relate heat flow and borehole temperatures with basement structure. Lower heat flow and geothermal gradients are correlated with thick sedimentary sequences over the basin, whereas high geothermal gradients depend on thinner sedimentary cover with a strong influence by available aquifers. Variations in the duration and temperature of fluid flow should affect the extent of mineral equilibration in the aquifers, with evidence for more complete re-equilibration over basement highs. This kinetic implication may be one reason why MVT mineralization is absent from dolomitized Palaeozoic carbonates in northwest Scotland, but present in the equivalent carbonates of Newfoundland. The Scottish carbonates are separated from the basement by a thick sequence of quartzo-feldspathic arenites whereas in Newfoundland there are no such underlying sediments. According to the new results, the geothermal gradient will have been locally higher in Newfoundland, which (coupled with fewer aquifers) enabled development of a coarse dolomite with pore space to host mineralization. In spite of extensive dolomitization in northwest Scotland, there is no such coarse crystalline dolomite.

A second aspect of heat transfer touched on by Deming *et al.* concerns the extent to which heat is transferred along the length of the fluid flow path. MVT deposits formed at the end of the fluid flow path have typical precipitation temperatures in the range 200–100 °C, raising questions on the heat budget of the fluid migration. In the North Slope system, the fluids are heated early in the deeper parts of the basin, and maintain their temperature of 200–100 °C for most of the fluid path. This observation removes superficial questions about the extent to which heat is transferred in the basin, but not about the underlying heat budget.

Finally Deming *et al.* provide a severe test of our current understanding of solution chemistry. In the North Slope, the average groundwater velocity is estimated to be 10 cm yr⁻¹ and, excepting climatic factors, there are no obvious reasons why these flow rates should not be applied elsewhere. Taking the Upper Mississippi Valley District as an example, it has been estimated by several independent techniques that a major ore body (10⁶ tonnes ZnS) takes 10⁵–10⁶

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 2 Thin section of sphalerite (ZnS) from Shullsburg ore body, Wisconsin. Variations in colours are due to variations in iron content over timescales of 10⁵–10⁶ years. The height of the section is 70 mm. (Photograph by S. Maher.)

years to form¹¹. A simplistic calculation at the temperatures and acidities inferred from the North Slope and other aquifers suggests the ore solution has to carry between 1–10 parts per million metals to form an ore body. This estimate is in agreement with other workers, and reinforces the opinion that metal chloride and bisulphide complexing alone are insufficient by approximately two orders of magnitude to account for the required solubility. The North Slope data imply that other metal complexes are required, perhaps organic ligands¹² which were derived during petroleum formation. □

Andrew P. Gize is in the Department of Geology, University of Manchester, Manchester M13 9PL, UK.

- Deming, D., Sass, J. H., Lachenbruch, A. H. & De Rito, R. F. *Geol. Soc. Amer. Bull.* **104**, 528–542 (1992).
- Cathles, L. M. & Smith, A. T. *Econ. Geol.* **78**, 983–1002 (1983).
- Garven, G. & Freeze, R. A. *Am. J. Sci.* **284**, 1085–1124 (1984).
- Evans, D. G. & Nunn, J. A. *J. geophys. Res.* **94**, 12,413–12,422 (1989).
- Lachenbruch, A. J. *et al.* *USGS Prof. Pap.* **1399**, 645–656 (1988).
- Sverjensky, D. A. & Garven, G. *Nature* **356**, 481–482 (1992).
- McLimens, R. K., Barnes, H. L. & Ohmoto, H. *Econ. Geol.* **75**, 351–361 (1980).
- Mason, S. E. thesis, Pennsylvania State Univ. (1987).
- Dutton, S. P. & Land, L. S. *Geol. Soc. Am. Bull.* **100**, 1271–1282 (1988).
- Burley, S. D., Mullis, J. & Matter, A. *Marine Petrol. Geol.* **6**, 97–120 (1989).
- Gize, A. P. & Barnes, H. L. *Econ. Geol.* **82**, 457–470 (1987).
- Giordano, T. H. & Barnes, H. L. *Econ. Geol.* **76**, 2200–2211 (1981).

The subunit story thickens

R. J. Lefkowitz

THE heterotrimeric G proteins¹, molecules that have a central role in signal transduction across the cell membrane, are composed of α , β and γ subunits. For several years it was the α subunit which held the limelight, with a great deal of attention being paid to its mediation of the activation of an ever-growing list of effector molecules such as enzymes and ion channels. The β and γ subunits, by contrast, were held to be the dull end of G proteins, serving merely to regulate the levels of the uncomplexed, active α subunit. That dogma has come increasingly into question through, for example, reports in *Science*² late in 1991 and in *Nature*³ earlier this year. Even more excitingly, we now have the report by Kleuss *et al.*, on page 424 of this issue⁴, which shows that specific β subunits are involved in coupling specific receptors to a calcium channel.

The G proteins were initially discovered as integral components of the adenylyl cyclase⁵ and retinal phototransduction⁶ systems, and were named after the characteristic property of their α -subunits, the ability to bind and hydrolyse guanine nucleotides. The other two subunits were discovered subsequently; they form a tightly associated complex, being dissociated only under denaturing conditions.

For all the attraction of the α -subunit, by 1989 prominent roles for $\beta\gamma$ in carrying receptor-derived information in several transmembrane signal transduction systems had begun to emerge (for example, pheromone-induced mating in *Saccharomyces cerevisiae* and, perhaps less convincingly, activation of PLA₂ (reviewed in ref. 7)). Experiments from two laboratories have now called into question what had seemed to be the very well-established solo role of α in activating adenylyl cyclase^{2,3}. Gilman and colleagues cloned several distinct forms of this enzyme (currently referred to as types I–IV (refs 8–11) with more reportedly on the way). Tang and Gilman² expressed several forms of these enzymes in SF9 cells by infection with recombinant baculovirus and then examined their *in vitro* regulation by α - and $\beta\gamma$ -subunits of G_s. Results with the type I adenylyl cyclase (which is a Ca²⁺/calmodulin-sensitive form) were pretty much as predicted — α_s stimulates and $\beta\gamma$ inhibits, presumably by complexing activated α_s because, by itself, $\beta\gamma$ has no effect. Results for the type II enzyme, however, were strikingly and surprisingly different. This form of the enzyme (which is Ca²⁺/calmodulin-insensitive) was stimulated to the same extent as the

type I by α_s , but was also strikingly stimulated by $\beta\gamma$ in the presence of activated α_s . Type III adenylyl cyclase showed relatively little sensitivity to $\beta\gamma$, and type IV was similar to type II in its properties¹¹.

Federman *et al.*³ essentially confirmed these findings, but they used a novel approach which immediately suggests the relevance of the new findings to receptor-mediated transmembrane signalling. When mammalian HEK293 cells were transfected with α_2 -adrenergic receptors (which generally inhibit adenylyl cyclase activity via G_i proteins), the receptors were found to mediate inhibition of synthesis of cyclic AMP in response to appropriate agonists. However, co-transfection with type II adenylyl cyclase converted this response to stimulation of cAMP synthesis. That this stimulation was indeed mediated by $\beta\gamma$ subunits of the G_i proteins was confirmed by its blockade following pertussis toxin treatment and by co-transfection with the α -subunit of transducin which 'scavenges' $\beta\gamma$ subunits but which cannot modulate adenylyl cyclase activity itself. In accord with the findings of Tang and Gilman², stimulation of cAMP synthesis by the G_i-coupled receptors required concurrent activation by α_s .

Taken together, these observations suggest that $\beta\gamma$ has a previously unappreciated signalling function, that of a 'conditional' activator of cAMP synthesis. But do they have any physiological relevance? Both groups point to previously confusing results in the literature which are rationalized by this new view of events. These results concern examples of agents which, while not themselves activators of adenylyl cyclase in brain slices (for example glutamate, α -adrenergic receptor agonists, GABA_B receptor agonists), are nonetheless able to potentiate compounds such as histamine, β -adrenergic agonists or adenosine¹². These effects, it is now proposed, are due to release of $\beta\gamma$ from G_i- or G_o-type G proteins which then stimulate a type II adenylyl cyclase.

As Federman *et al.* say, the presence of multiple effector adenylyl cyclases significantly extends the range and versatility of G-protein-coupled receptor signalling and opens a variety of lines of 'cross-talk' between such systems. But if essentially all G-protein-coupled receptors are capable of modulating the activity of the adenylyl cyclase pathway in a particular cell containing the type II enzyme, requisite physiological precision and specificity might be lost. This raises

the provocative issue of whether there are any specificity constraints built in.

Although the number of G-protein α -subunits discovered to date by molecular cloning stands at about 20, it is only quite recently that the diversity of β and γ subunits has been appreciated (four forms of each cloned thus far). Might they differ in their functional attributes? It is in this context that the paper by Kleuss *et al.*⁴ is particularly relevant and intriguing. These investigators used somatostatin and muscarinic cholinergic receptor-mediated inhibition of Ca²⁺ currents in rat pituitary GH₃ cells, and nuclear microinjection of short, selective antisense oligonucleotides directed at four distinct G-protein β subunits, to clarify whether there is differential involvement of the subunits in coupling the two different receptors to the same effector system. Their data indicate that β_1 is involved in somatostatin receptor signalling, β_3 in muscarinic cholinergic receptor signalling and β_2 and β_4 in neither. Thus, different $\beta\gamma$ subunits seem to be contained within the heterotrimeric G proteins interacting with these two G_o-coupled receptors. The contribution, if any, of different γ subunits remains to be seen.

If stimulation of different receptors leads to the release of distinct populations of $\beta\gamma$, then there is the potential for receptor specificity in the effector functions of $\beta\gamma$ such as stimulation of the type II adenylyl cyclase. This would simply require that the different $\beta\gamma$ isoforms have distinct functional properties with respect to effector stimulation. Whether there are indeed such differences in the ability of different $\beta\gamma$ isoforms to activate adenylyl cyclase is a hot question. Equally provocative are such issues as whether other effectors are regulated by $\beta\gamma$ (for example various phospholipases) and what the underlying molecular mechanisms of such effects might be. □

Robert J. Lefkowitz is in the Howard Hughes Medical Institute Research Laboratories, Duke University Medical Center, Durham, North Carolina 27710, USA.

1. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
2. Tang, W. J. & Gilman, A. G. *Science* **254**, 1500–1503 (1991).
3. Federman, A. D., Konkin, B. R., Schrader, K. A., Reed, R. R. & Bourne, H. R. *Nature* **356**, 159–161 (1992).
4. Kleuss, C., Scherübel, H., Hescheler, J., Schultz, G. & Wittig, B. *Nature* **358**, 424–426 (1992).
5. Gilman, A. G. *Rev. Biochem.* **56**, 615–649 (1987).
6. Stryer, L. & Bourne, H. R. *Rev. Cell Biol.* **2**, 391–419 (1986).
7. Bourne, H. R. *Nature* **337**, 504–505 (1989).
8. Krupinski, J. *et al. Science* **244**, 1558–1564 (1989).
9. Feinstein, P. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10173–10177 (1991).
10. Bakalyar, H. A. & Reed, R. R. *Science* **250**, 1403–1406 (1990).
11. Gao, B. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10178–10182 (1992).
12. Sattin, A., Rall, T. W. & Zanella, J. J. *Pharmac. exp. Ther.* **192**, 22–32 (1975).

The structure of smart fluids

M. Whittle and W. A. Bullough

ELECTRO-RHEOLOGICAL fluids, which can be transformed into a gel-like state by applying an electric field, are the subject of increasing interest in engineering sciences, yet little is known of the microscopic basis of their remarkable properties. Tao and colleagues, at the Southern Illinois University, have devised a laser-diffraction technique for revealing the microscopic structures of such fluids¹, and confirm the arrangement of the electrified fluids predicted by their models.

Among the most spectacular of 'smart' materials, electro-rheological fluids are slurries of particles suspended in an insulating fluid. On application of a strong electric field, perhaps 1 megavolt per metre, they become plastic in the sense of exhibiting a yield stress. The rheological changes are associated with the formation of chains and fibrils of the suspended particles under the influence of the applied field². But it is hard to see what is going on at the volume fractions (20–40 per cent) with which useful rheological changes occur, because of the opacity of the fluids. Instead, observations are typically made with dilute solutions (around 2 per cent volume fraction), in which the chains can be seen to coalesce into columns (see figure) and even (at higher concentrations) into extended three-dimensional structures.

The mechanism of this behaviour is normally discussed in terms of particle polarization^{2,3} although this is not yet quantitatively established as the only mechanism and other effects may be present. In earlier work, Tao and Sun⁴ predicted a body-centred tetragonal (b.c.t.) lattice structure for the columns on the basis of energy minimization for a system of induced dipoles. This is related to the better known body-centred cubic but only two sides of the unit cell are equivalent in the primitive lattice.

Tao and colleagues have now devised a novel laser-diffraction method which confirms that the column structure is body-centred cubic at a volume fraction of 20 per cent. At these concentrations light scattering is useless because of the high degree of multiple scattering. The new method relies on the use of a suspension of glass microspheres of highly uniform diameter. Although glass spheres do not give very good electro-rheological fluids (in terms of the yield stress achievable), we may expect them to produce representative structures.

The resulting diffraction pattern for the fluid subjected to a field is not related to any conventional diffraction

pattern and the authors propose that it is the result of multiple lensing by particles arranged in a regular lattice. A textbook analysis of a linear periodic array of focusing elements reveals that light can propagate along the lattice under certain circumstances and allows the prediction of the pattern imposed on the exit beam. The authors find that only the b.c.t. structure can account for the observed pattern.

There has always been considerable

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Columns of dilute cornstarch in motor oil (5 per cent by mass) in a field of 2.5 megavolts per metre. Depth of figure, 365 μm . (Courtesy of D. G. Froode.)

argument about the nature of the interaction in electro-rheological fluids and the fact that the predicted structure is revealed experimentally supports the case for dipole–dipole interactions. Nevertheless, it leaves ample scope for discussion as to how the dipoles arise in the first place — indeed this may differ from fluid to fluid. The nature of the structure also has important predictive potential, as the relation between physical properties and extended structure is well documented. Transport properties such as thermal and electrical conductivity follow the lattice symmetry and will therefore exhibit anisotropy. Mechanical properties like the yield stress will also be related and should exhibit preferential slip planes. Recent work into the use of relatively dilute fluids for display systems and anti-radiation films could directly benefit from an understanding of the anisotropy, as these devices operate under static conditions and use relatively dilute suspensions. In this case, as in many applications, a rapid response of the fluid is crucial and the dynamic aspect of structure formation is an area that this work could now extend to.

Any comparison between theory and experiment is welcome in this respect and it resolves an old question. But much remains to be done. There is a gulf between these kinds of experiments and the real world of engineering requirements. These are complicated systems and many experiments and theoretical

approaches are understandably made far from the regime where electro-rheological fluids are actually useful. The real fluids are composed of polydisperse irregularly shaped particles unlikely to settle into a regular lattice structure and there is no evidence as yet to suggest that the extra effort and cost involved in producing monodisperse systems will improve their useful properties.

In this context the stability of particles under hostile engineering conditions must also be considered. Electro-rheological fluids are currently under investigation with a view to providing rapid coupling between components in high-speed machines (the gel-like state can be induced within hundreds of microseconds), and in these applications they are subjected to a shear field in addition to the electric field. The competitive edge offered by these devices is in the speed of response and this is strongly dependent upon particle concentration. Generally, volume fractions of 30 per cent upwards are necessary to produce viable devices. The response time is further influenced by the shear rate and temperature.

The shear rate in an operating electro-rheological valve may be as high as $20,000\text{ s}^{-1}$, and for clutches centrifugal forces of around $300g$ may arise⁵. These conditions can cause considerable practical problems in the guise of sedimentation or plating out of the particle phase. Computer simulations suggest that even at relatively low shear rates the particles arrange themselves into sheets parallel to the flow^{6,7}. It seems unlikely that such dynamic structures can be described with the simplicity that arises at zero shear rate, but the easier theoretical situation is related to how quickly the dynamic equilibrium is attained and decays in response to a changing applied field.

It is important to ask answerable questions and to give answers as excellent as Tao and colleagues have done. But to tackle the wider commercially important problems needs collaboration between workers who have access to materials properties, engineering-scale experimental data on time response and computer simulations. □

M. Whittle and W. A. Bullough are in the Department of Mechanical Engineering and Process Control, University of Sheffield, Sheffield S1 3JD, UK.

1. Chen, T., Zitter, R. N. & Tao, R. *Phys. Rev. Lett.* **68**, 2555–2558 (1992).
2. Tao, R. (ed.) *Electrorheological Fluids* (World Scientific, Singapore, 1992).
3. Block, H. & Kelly, J. P. *J. Phys. D* **21**, 1661–1677 (1988).
4. Tao, R. & Sun, J. M. *Phys. Rev. Lett.* **67**, 398–401 (1991).
5. Bullough, W. A. et al. *Mechatronic Syst. Eng.* (in the press).
6. Whittle, M. J. *Non-Newtonian Fluid Mech.* **37**, 233–263 (1990).
7. Melrose, J. *Molec. Phys.* (in the press).

A phantom of the ocean

Theodore J. Smayda

A CHARMING explanation of so-called 'red tides' is to be found in the rich folklore of the sea: the characteristic phosphorescence (bioluminescence) and associated surface discolorations are, it is said, produced by an enormous fish, The Devil of the Waters, which inhabits the seabed, spitting fire to destroy its prey, and thus reddening the waters¹. Well, this fictitious monster would have met its match in the marine David described by Burkholder and her colleagues on page 407 of this issue². They report the discovery of a microscopic, unicellular organism which, at one stage in its remarkable life cycle, secretes a powerful toxin which kills fish; most unusually, the organism uses dying fish to its nutritional advantage.

The organism concerned is a photosynthetic dinoflagellate, and it lurks dormant on the seabed until live fish approach. Then, sometimes within minutes, it excysts, releasing a motile, vegetative stage which swims, grows and swarms within the water column. There it secretes a potent, water-soluble neurotoxin which causes fish death, sometimes on a massive scale — in one episode, about a million menhaden (*Brevoortia* sp.) were killed. Several hours later, the dinoflagellate cells encyst, sink to the bottom sediments and await a renewed, ichthyo-stimulated foray.

The authors appropriately liken this stunningly rapid and ephemeral sequence of lethal appearance and disappearance to that of a phantom. The chemical nature of the neurotoxin remains to be elucidated, as does the broader distribution of the dinoflagellate. But although it is known only from two estuaries in the United States, the authors predict that it is widespread and has been responsible for many of the enigmatic fish kills which occur in coastal regions.

Red tides (which may in fact be red, green, yellow or brown) are of course one manifestation of blooms of phytoplankton, population explosions undergone by dinoflagellates, diatoms and several other groups of photosynthetic microalgae. Such blooms have underpinned marine foodwebs for at least three billion years, but the red-tide blooms of nutritionally inadequate or toxic species have less benign effects. That has been known for a long time. What is new is an apparent increase in bloom frequency and spread of noxious species; the occurrence of toxicity in species that had been thought to be harmless; and the growing number of reports of catastrophic mass deaths of marine mammals, including whales³, and of fish and

invertebrates⁴.

Nor can the public at large stand aloof from these phenomena. In the best known of the toxic bloom cases, paralytic shellfish poisons' (PSP) toxicity, shellfish feeding upon certain dinoflagellates accumulate the neurotoxic saxitoxin and its analogues without harming themselves. But when the shellfish are eaten by humans, serious illness or death can occur: a recent PSP outbreak



The "phantom-like" dinoflagellate described by Burkholder *et al.*². The toxic vegetative cells extend a tongue-like organelle, the peduncle (shorter structure, bottom), which attaches to and sucks the contents from fragments of fish tissue.

in Guatemala resulted in 26 deaths and 187 cases of illness⁵. Likewise, reports of fish deaths caused by red-tide outbreaks are not new as such^{4,6,7}. But we are seeing the emergence of hitherto unsuspected toxins and toxic syndromes, such as the 1987 discovery of the amnesic shellfish poisoning in Cardigan Bay, Nova Scotia, caused by the neuroexcitatory amino acid, domoic acid⁸. Before then, toxin production by diatoms, a major component of the phytoplankton, was unknown. This toxin also suddenly appeared in Monterey Bay, California, in 1991, where it caused the death of pelicans feeding on anchovy which had in turn accumulated domoic acid from its diatom prey⁹.

Another recent development, underlining how easily the phytoplankton can spring surprises on us, was the report of a previously undescribed mode of feeding (termed dasmotrophy) by a toxic, photosynthetic phytoplankter on other phytoplankton¹⁰. This toxic flagellate, also lethal to fish, invertebrates and macroalgae, is thought to produce

membrane-puncturing compounds which allow it to extract nutrients from its prey.

It is against this background that the paper of Burkholder *et al.* must be viewed. With the possible exception of dasmotrophy, phytoplankton toxins have not been reported to have a nutritional role; rather, it is thought that they may serve as chemical deterrents against predation, even protecting shellfish which have sequestered them during feeding¹¹. By contrast, the dinoflagellate described by Burkholder *et al.* does derive nutritional gain during its kill — it digests flecks of sloughed-off fish tissue to which it attaches by means of a peduncle. Another difference between the biology of this dinoflagellate and that of other toxin producers is in the surprising influence of phosphorus. The addition of phosphate, unlike nitrate or ammonia, stimulated the growth of gametes, whereas it is limitation of this nutrient that has been implicated in toxigenesis in other species¹².

Burkholder *et al.* point out that the apparent correlation between increasing nutrient enrichment of global coastal waters and increased incidences of harmful algal blooms is probably not a factor in blooms of the phantom dinoflagellate. But a sobering consequence of their discovery is that it further complicates resolution of the matter of whether harmful algal blooms in the sea are actually increasing, or whether the increase is due to improved monitoring of such events. However that may be, the public are subject to periodic alarms over the safety of seafood, and fisheries and aquaculture enterprises suffer considerable economic loss.

Several big questions need to be resolved. Are toxic blooms in general being triggered largely by increased nut-

1. Morvan, F. *Legends of the Sea* (Crown, New York, 1980).
2. Burkholder, J. M., Noga, E. J., Hobbs, C. H. & Glasgow, H. B. *Jr Nature* **358**, 407–410 (1992).
3. Geraci, J. R. *et al. Can. J. Fish. Aquat. Sci.* **46**, 1895–1898 (1989).
4. Smayda, T. J. in *Food Chains: Yields, Models and Management of Large Marine Ecosystems* (eds Sherman, K., Alexander, L. M. & Gold, B. D.) 275–307 (Westview, Boulder, 1989).
5. Rosales-Loessener, F., De Porras, E. & Dix, D. W. in *Red Tides: Biology, Environmental Science and Toxicology* (eds Okaichi, T., Anderson, D. M. & Nemoto, T.) 113–116 (Elsevier, New York, 1989).
6. Roberts, B. S. in *Toxic Dinoflagellate Blooms* (eds Taylor, D. L. & Seliger, H. H.) 199–202 (Elsevier/North Holland, New York, 1979).
7. *Toxic Marine Phytoplankton* (eds Granéli, E., Sundström, B., Edler, L. & Anderson, D. M.) 1–554 (Elsevier, New York, 1991).
8. Bates, S. S. *et al. Can. J. Fish. Aquat. Sci.* **46**, 1203–1215 (1989).
9. Work, T. M. *et al. in Toxic Phytoplankton Blooms in the Sea* (eds Smayda, T. J. & Shimizu, Y.) (Elsevier, Amsterdam, in the press).
10. Estep, K. G. & MacIntyre, F. *Mar. Ecol. Prog. Ser.* **57**, 11–21 (1989).
11. Kvitsek, R. G., De Grange, A. R. & Beittler, M. K. *Limnol. Ocean.* **36**, 393–404 (1991).
12. Edvardsen, B., Moy, F. & Paasche, E. in *Toxic Marine Phytoplankton* (eds Granéli, E., Sundström, B., Edler, L. & Anderson, D. M.) 284–289 (Elsevier, New York, 1991).

rient loadings in coastal waters, or by fisheries and aquacultural activities, or are they manifestations of long-term cyclical trends? Are they perhaps a worrying consequence of global deterioration of the marine environment? The next chance for all concerned to put their heads together and discuss these issues will be in October 1993, when the

sixth international conference on toxic blooms takes place. We cannot hope to have the answers by then, but we can hope for more data to go on. □

Theodore J. Smayda is at the Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA.

TRANSCRIPTIONAL REGULATION

A closer look at E2F

Joseph R. Nevins

ADENOVIRUS, a mammalian DNA tumour virus, has been a key player in investigations of fundamental cellular control mechanisms, as the virus must subvert the cell's machinery for transcription and replication for its own purposes. In 1986, studies of transcription control mediated by the adenovirus protein E1A identified a cellular DNA-binding protein, termed E2F, that recognizes key regulatory elements in the viral E2 gene promoter¹. Further experiments defined the role of the factor in the regulation of transcription of the viral gene, but the significance of this cellular protein beyond the realm of adenovirus was unclear.

Things have changed over the past year, and dramatically — E2F has been thrust into the midst of cellular proliferation control mechanisms that are attacked by viral oncoproteins. This intense activity has now reached a crescendo with the cloning of the gene encoding E2F, as reported by Helin *et al.*² and Kaelin *et al.*³ in the latest issue of *Cell*. A protein which had previously been studied as a band in a gel retardation assay is now ready for a closer look.

Oncogenic action

The recent developments of the E2F story actually began in 1988, involving studies of the oncogenic action of E1A, when it was realized that one of the cellular proteins originally identified as an E1A-binding protein, and thus a candidate target for E1A action, was the product of the retinoblastoma (Rb) gene⁴. It soon emerged that each of the DNA tumour viruses, including adenovirus, SV40 and human papillomavirus, encode proteins that bind the Rb protein. Clearly, these otherwise distinct viruses had a common activity, presumably essential for viral replication, that intersected with the action of a cellular tumour suppressor protein. But what was the activity and what function of the Rb protein was altered?

Analysis of the effect of viral E1A on the activity of E2F provided important clues. In adenovirus-infected cells, E2F

is complexed to a 19K product of the viral E4 gene, an interaction that allows E2F to bind in a cooperative fashion to the E2 promoter. It turns out that the formation of the E2F–E4 complex requires the assistance of the E1A gene product because, in most cell types, E2F is complexed with cellular proteins that prevent its interaction with the E4 protein. An *in vitro* assay demonstrated that E1A could release E2F from these cellular protein complexes, making it available to bind to the E4 protein and hence the E2 gene promoter⁵.

When it became clear that the same E1A domains were needed for the dissociation of these E2F complexes as were required for E1A binding to proteins such as Rb, a unifying mechanism became evident (see figure). That is, if Rb were one of the proteins associated with E2F, then the process of E1A-mediated dissociation of the complex might transfer Rb to E1A. Analysis of the E2F complexes for the presence of the Rb protein showed that this was indeed the case^{6–8}. An independent approach demonstrated that complexes of Rb with cellular proteins possess sequence-specific DNA-binding activity specific for the E2F recognition sequence⁹. Thus, two approaches reached the same conclusion — E2F and Rb are cellular partners.

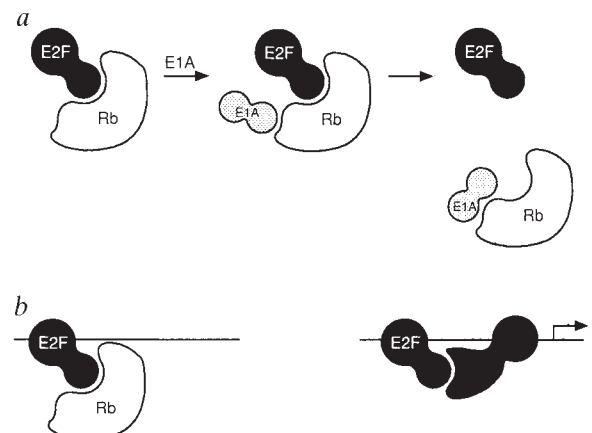
It was the ability of the Rb protein to interact specifically with E2F that has led to the cloning of a complementary DNA encoding E2F. Expression libraries initially probed with labelled Rb protein yielded several clones¹⁰, but none of them turned out to be E2F. Additional screening, however, brought to light clones termed *RBP3* by Helin *et al.*² and *RBAP-1* by Kaelin *et al.*³, which encode iden-

tical proteins with properties suggestive of E2F. The molecular weight of the encoded protein is close to that previously reported for E2F and the expressed protein exhibits the appropriate E2F DNA-binding properties. Antibodies specific for the cloned protein immunoprecipitate E2F activity and detect a protein of appropriate molecular weight in an enriched preparation of E2F. Moreover, the cloned protein interacts with the adenovirus E4 protein, as well as with the Rb protein, but not with mutant versions of these proteins. Finally, expression of the cDNA in transfected cells stimulated E2F-dependent transcription of a reporter gene.

Strong evidence

Taken together, then, the data in the two papers provide strong evidence that the clone does indeed encode E2F. But one question to be resolved is whether it encodes all of the E2F activity in the cell. Attempts to abolish or alter the mobility of the family of E2F complexes that are detected in cell extracts met with only partial success. Although technical difficulties could be the explanation, there may be more to come.

What does the sequence of the E2F clone tell us about its function? Although comparison with known sequences offers no evidence that E2F has been previously cloned, or that it is a close relative of a known protein, a 102-amino-acid segment essential for DNA binding contains a basic helix-loop-helix motif. It is not yet known if this motif is essential for the DNA-binding capacity of E2F, but it does raise the possibility that E2F functions as a protein dimer, or even as a heterodimer, possibly changing the DNA-binding specificity and thus generating an entirely



a, The E2F–Rb complex is depicted as an association of Rb with the carboxy-terminal domain of E2F. As a consequence of E1A action, E2F is displaced from the complex, leaving E1A bound to Rb. The intermediate involving an association of E1A with the E2F–Rb complex is speculative. b, The interaction of E2F with a promoter and the possible influence of Rb on the contacts normally made with other components of the transcriptional apparatus.

new set of target genes.

A relatively small domain of E2F, consisting of 18 amino acids that reside at the extreme carboxy terminus of the protein (residues 409–426), seems to be responsible for the interaction with Rb. The Rb-binding domain also overlaps with sequences that are essential for transcriptional activation, which represent the carboxy-terminal 60 amino acids, at least when assayed in GAL4 fusion constructs. Given this overlap, a simple mechanism for Rb control of E2F function could be envisaged whereby the Rb protein would mask the activation domain, blocking the ability of E2F to interact with the transcriptional machinery and stimulate transcription (see figure). In fact, transfection experiments have shown that the interaction of Rb with E2F, although having no effect on DNA binding, inhibits the ability of E2F to stimulate transcription¹¹.

Common pathway

This brings us back to the question of why E2F, and the interaction of E2F with proteins such as Rb, is sitting at the centre of a common pathway targeted by the DNA virus oncoproteins. One possible answer comes from the identification of cellular genes that appear to use E2F. Many of them encode proteins that are essential for S phase in the cell cycle, including dihydrofolate reductase (DHFR), thymidine kinase and DNA polymerase α . Moreover, the E2F sites in the DHFR promoter are essential for activation of DHFR in late G1 (ref. 12). Uncomplexed, and presumably transcriptionally active, E2F accumulates at this time¹³. So a common strategy of the DNA tumour viruses may be to activate E2F, and thus S-phase gene expression, in otherwise quiescent cells to provide the necessary environment for viral DNA synthesis.

Simple stories usually become more complicated before they are finished. The finding⁷ that only the underphosphorylated form of Rb was bound to E2F, an observation now also made with the cloned E2F protein, led to the speculation that E2F was inactive in G1, as a consequence of sequestration by Rb, and was then released as an active factor in late G1 when Rb became phosphorylated. Although uncomplexed E2F does accumulate at the end of G1, release of the E2F–Rb complex does not appear to be the source because the E2F interaction with Rb persists through S phase (ref. 14 and unpublished observations of my own group). There is a G1-specific E2F complex¹⁵ that could be a source of this E2F, but this complex does not involve Rb. New synthesis of E2F might also contribute under some circumstances, as Kaelin *et al.*³ found that the RNA detected by the *RBAP-1* cDNA

was at low levels in quiescent T cells and then accumulated following mitogen stimulation.

But what, then, is the functional significance of the Rb–E2F interaction if it does not serve as a reservoir for E2F? One possibility is that Rb bound to E2F does not simply sequester E2F, but that this complex actively represses the transcription of certain target genes. Indeed, experiments reported earlier this month in *Nature*¹⁵ point to this possibility. Two potential targets for control by E2F–Rb are the *c-myc* gene and the *cdc2* gene — transfection assays demonstrate that Rb can repress *c-myc* transcription as well as *cdc2* transcription, dependent on E2F sites in the respective promoters^{16,17}.

Finally, as if the array of interactions involving E2F were not complicated enough, E2F is also found in association with the Rb-related p107 protein, together with the cyclin A–cdk2 kinase complex, during S phase^{14,18,19}. This association thus generates a cyclin-dependent protein kinase that has sequence-specific DNA-binding activity, an interesting property for a kinase that is suspected to be involved in the events of S phase. Although it seems unlikely that the complex would be essential for the initiation of S phase, given that transformed cells lacking the complex certainly do enter S phase, a role at the end of S phase in controlling the entry into mitosis is a possibility. Clearly, one payoff of the cloning of E2F will be the generation of reagents that should provide answers to questions as to the precise function of E2F in these events. Regardless of the details, it is already clear that this transcription factor is a central figure in the control of cell proliferation, gaining the attention of a variety of cellular and viral regulatory proteins — not to mention a few scientists. □

Joseph R. Nevins is in the Section of Genetics, Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA.

- Kovesdi, I., Reichel, R. & Nevins, J. R. *Cell* **45**, 219–228 (1986).
- Helin, K. *et al.* *Cell* **70**, 337–350 (1992).
- Kaelin, W. G. *et al.* *Cell* **70**, 351–364 (1992).
- Whyte, P. *et al.* *Nature* **334**, 124–129 (1988).
- Bagchi, S., Raychaudhuri, P. & Nevins, J. R. *Cell* **62**, 659–669 (1990).
- Bandara, L. R. & LaThangue, N. B. *Nature* **351**, 494–497 (1991).
- Chittenden, S. P. *et al.* *Cell* **65**, 1053–1061 (1991).
- Bagchi, S., Weinmann, R. & Raychaudhuri, P. *Cell* **65**, 1063–1072 (1991).
- Chittenden, S. P. *et al.* *Cell* **65**, 1073–1082 (1991).
- DeFeo-Jones, D. *et al.* *Nature* **352**, 251–254 (1991).
- Hiebert, S. W. *et al.* *Genes Dev.* **6**, 177–185 (1992).
- Means, A. L. *et al.* *Molec. cell. Biol.* **12**, 1054–1063 (1992).
- Mudryj, M. *et al.* *Cell* **65**, 1243–1253 (1991).
- Shirodkar, S. *et al.* *Cell* **68**, 157–166 (1992).
- Weintraub, S. J., Prater, C. A. & Dean, D. C. *Nature* **358**, 259–261 (1992).
- Hamel, P. A. *et al.* *Molec. cell. Biol.* (in the press).
- Dalton, S. *EMBO J.* **11**, 1797–1804 (1992).
- Devoto, S. H. *et al.* *Cell* **68**, 167–176 (1992).
- Cao, L. *et al.* *Nature* **355**, 176–179 (1992).

DAEDALUS

Wood for ever!

A TREE, says Daedalus, is only alive on the outside. Every year it grows another circumferential surface layer of living wood, while the layer beneath dies. The central wooden core is only a passive support; even if it rots, the resulting hollow tree can continue to flourish. So it should be possible to remove a hollow core from a tree every year, and trust to the inward pressure of next year's growth to squash the central hole back to its original size, ready to yield another hollow core.

This subtle concept of 'continuous forestry' deserves to be elaborated. Daedalus plans to plant a line of trees close together, like close-packed railings in a fence. He will chop off all branches which grow in the plane of this fence; as the trunks themselves grow and make mutual contact, he will shave the bark along the contact line and join the living wood beneath with grafting wax. The line of trees should ultimately graft completely together, fusing into a sort of 'planar tree' with branches sticking out on either side of it.

The next stage of the operation will be rather tricky. The planar tree will be sawn vertically down its long axis, separating it into two half planes. Each will have bark and branches on its front face, while its raw wooden back face will abut that of its twin. One half plane will be uprooted, moved a few metres back, and replanted. The resulting parallel wooden walls will resemble the front and back walls of a house. The enclosure will be completed by transplanting a second pair of tree half planes into place as side walls, and adding a suitable greenhouse-type roof.

The result will be a sort of huge square hollow tree-house with a sheltered interior. New wood will be steadily laid down on the living outer surface, and will displace the inner dead layers into the house. The warm, dry atmosphere within will season them to perfection. Every few years or so, a thick layer will be planed off the inner walls of the house and manoeuvred out of a vertical exit slit left in one corner. An endless supply of huge, ready-seasoned wooden sheets will be obtained, far larger than the narrow rectangular planks yielded by chopping down a natural tree, and without the waste imposed by its circular cross-section.

Trees have no natural lifespan; once established, an estate of tree-houses should survive indefinitely. Nothing need ever be cut down. The clusters of living, branched, leafy, but recognizably rectangular houses will have a fantastic, magical look, like something out of Hans Andersen. They could even be let as dwellings. The ideal tenants would be witches, lost children and enchanted princesses.

David Jones

Molecular mimicry in liver disease

SIR — Primary biliary cirrhosis (PBC), an autoimmune chronic cholestatic liver disease, affects mainly women and ultimately leads to the destruction of the intra-hepatic bile ducts and death from liver failure. The disease is characterized by the presence of diagnostic serum anti-mitochondrial antibodies¹.

We have found low titres of these antibodies in more than 60% of women with recurrent urinary tract infections but not suffering from liver disease, and find that women with PBC are more susceptible to recurrent urinary tract infections than those with other types of liver disease². The bile ducts in PBC abnormally express class II HLA-DR and HLA-DQ antigens³, and are infiltrated by lymphocytes, predominantly CD4⁺ and CD8⁺. We now propose that the autoimmune phenomena result from T-cell epitopes of the microbial proteins being mimicked by peptides derived from, and presented by, the abnormally expressed HLA-DR antigen (Fig. 1).

The targets of the mitochondrial autoantibodies have been identified¹. The

major antigen (M2) contains at least six distinct polypeptides, three of which are the E2 components (dihydrolipoamide acyltransferases) of the related multi-enzyme complexes: pyruvate dehydrogenase (PDC), branched-chain oxoacid dehydrogenase complex (BCOADC) and oxoglutarate dehydrogenase complex (OGDC). The others are the lipoyl-containing protein X and the E1- α and - β polypeptides of PDC. Anti-E2 antibody recognizes a component of the outer membrane surface of biliary epithelial cells⁴. Two other PBC-specific autoantibodies recognize antigens termed M4 (sulphite oxidase) and M9 (glycogen phosphorylase)¹. Little is known about T-cell epitopes; T cells have been cloned, but only some respond to PDC-E2 (ref. 5).

E2 peptides are evolutionarily highly conserved and antibodies from patients with PBC react with both human PDC-E2 and *Escherichia coli* PDC-E2 (ref. 6). Although the respective antibodies are largely peptide-specific, studies have shown that the chief linear epitope in all

E2s is around the residues DKA, where K is attached to the lipoyl moiety. Because of the similarities between human and *E. coli* PDC-E2, we explored this DKA region for similarities with other proteins, including HLA antigens.

Figure 2 shows that a 17-residue section (80–96) of HLA-DR- α has seven residues in common with the lipoyl regions I and II of *E. coli* PDC-E2 and six residues with lipoyl region III. There are as many similarities between this section of HLA-DR- α and corresponding sequences of the respective E2s as there are between the E2s of the different enzymes. The DKA sequence was also present in the human muscle glycogen phosphorylase (M9) and chicken sulphite oxidase (M4). But these sequence similarities cannot generate a common epitope to an antibody as the three-dimensional structure of the lipoyl region is a turn between β -strands⁷, whereas the segment in HLA-DR is α -helical⁸.

Instead, we propose that molecular mimicry occurs when these peptides bind as processed fragments to a histocompatibility antigen for recognition by T cells. Peptide ASFEAQGALANIADVKA from I-E α -chain (the mouse equivalent of HLA-DR) has recently been shown to bind to MHC I-A (equivalent to HLA-DQ)⁹. This implies that in humans the corresponding region of HLA-DR will be a T-cell epitope, and we suggest that at least one of the lipoyl regions of PDC-E2 also binds HLA.

Our hypothesis is that, in response to an infection, T cells recognize the lipoyl region of *E. coli* PDC-E2 and other similar enzymes; that these T cells stimulate a conventional antibody response to the infecting immunogens; that molecular mimicry subsequently occurs when these T cells recognize a self peptide (from HLA-DR- α), presented by aberrantly expressed class II HLA on biliary epithelial cells; and finally, that this initiates an autoimmune cascade leading to destruction of bile ducts in which mitochondrial autoantibodies and/or T cells may be pathogenetic.

Although the sequence similarities in Fig. 2 may be fortuitous, as may be the fact that the I-E α -peptide is found in

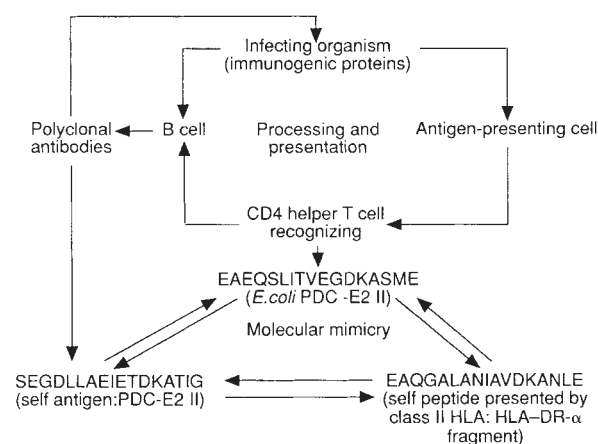


FIG. 1 Hypothesis of molecular mimicry between self and foreign T-cell epitopes in PBC.

FIG. 2 Identities of lipoyl regions with HLA-DR- α . Bold, residues common between HLA-DR- α and the other proteins; asterisks, residues common between HLA-DR- α and *E. coli* PDC-E2 I and II; cross, number of residues in common with HLA-DR- α ; open circle, number of residues in common with *E. coli* (PDC-E2 I). To gauge the significance of this observation, the 17-residue sequence

HLA-DR- α	80	E A Q G A L A N I A V D K A N L E	+	07
<i>E.coli</i> PDC-E2 I	29	E A E Q S L I T V E G D K A S M E		7
<i>E.coli</i> PDC-E2 II	132	E A E Q S L I T V E G D K A S M E		7
<i>E.coli</i> PDC-E2 III	233	A A E Q S L I T V E G D K A S M E		6
Human PDC-E2 I	88	N E G D L I A E V E T D K A T V G		45
Human PDC-E2 II	215	S E G D L L A E I E T D K A T I G		65
<i>S.cerevisiae</i> OGDC-E2	32	A E D Q V V A D V M T D K A T V E		56
<i>P.putida</i> BCOADC	102	K E D E L L A T I E T D K I D I E		67
Human glycogen phosphorylase	352	R I L V D L E R M D W D K A W D V		44
Sulphite oxidase	390	G R T W K V A R L M G D K A F P G		44

(29–45) for region I of *E. coli* PDC-E2 was scanned against the release 26 of the Protein Information Resource database for matches. We retrieved all 17-residue peptides that had an exact match for DKA and had another four or more identical residues in any position. Out of 20,354 database sequences, the matches were with HLA-DR, *E. coli* PDC-E2, *Azotobacter vinelandii* PDC-E2 and five other proteins (I-E- α was not in the database). Of these five other proteins, two were human (the insulin receptor and DNA-binding protein TR2). This shows that the sequences of these regions of *E. coli* PDC-E2 and HLA-DR- α chain are unusually similar. We are not suggesting that they are evolutionarily related.

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

association with MHC I-A, we suggest that it offers an explanation for the autoimmune basis of PBC. Similar mechanisms could occur in other autoimmune diseases, particularly in those in which autoantigens contain a similar motif to that identified here, and when there is abnormal expression of HLA in the target tissue, as for example in juvenile onset diabetes.

Andrew K. Burroughs

Patrice Butler

Royal Free Hospital and School
of Medicine,

Pond Street,

London NW3 2QG, UK

Michael J. E. Sternberg

Imperial Cancer Research Fund,

PO Box 123, 44 Lincoln's Inn Fields,

London WC2A 3PX, UK

Harold Baum

King's College,

Camden Hill Road,

London W8 7AH, UK

1. Bassendine, M. F. & Yeaman, S. J. *Hepatology* **15**, 545–548 (1992).
2. Butler, P. et al. *J. Hepatol.* (in the press).
3. Ballardini, G. et al. *Lancet* **II**, 1009–1013 (1984).
4. Joplin, R., Gordon Lindsay, J., Johnson, G., Strain, A. & Neuberger, J. *Lancet* **339**, 93–94 (1992).
5. Van de Water, J. et al. *J. Immun.* **146**, 89–84 (1991).
6. Lindenborn-Fotinos, J., Baum, H. & Berg, P. A. *Hepatology* **5**, 763–769 (1985).
7. Dardel, F., Laue, E. D. & Perham, R. N. *Eur. J. Biochem.* **201**, 203–209 (1991).
8. Brown, J. H. et al. *Nature* **332**, 845–850 (1988).
9. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S. C., Barlow, A. & Janeway, C. A. *Nature* **353**, 622–627 (1991).

Marine isotope evolution

SIR — McArthur *et al.*¹ criticized aspects of our study of Maastrichtian seawater ⁸⁷Sr/⁸⁶Sr evolution². Their primary contention is that our calculated rates of marine Sr isotope evolution are incorrect because of error in locating the Upper–Lower Maastrichtian boundary, and because of unrecognized faulting of the section. Both these criticisms are incorrect. Further, their Scientific Correspondence does not refute the fundamental observations of our study — that a break in slope in the strontium curve occurs well below the level of the Cretaceous/Tertiary (K/T) boundary and that the previously identified $1-2 \times 10^{-4}$ Sr isotope anomaly does not occur at this boundary. Their attempt to explain away our observations is flawed by misrepresentation of our data and is contradicted by their own published palaeontological data.

McArthur *et al.* claimed that Upper Maastrichtian strata do not occur until our 70-m horizon based on nannofossils and ammonites, while we identify it at 105–110 m. A recent zonation³ (based on nannofossil data obtained from the

same samples that we analysed for Sr isotopes) by one of the authors of ref. 1 places the Lower/Upper Maastrichtian between 95 and 150 m.

The contention that an “Early Maastrichtian age for this section immediately below 70 m is supported by the occurrence of *Pachydiscus neubergicus*” is outdated. It is now known that subspecies of this taxon range through most of the Maastrichtian^{4,5}. The taxon in question at the 70-m locality is, in fact, *P. neubergicus dissitus* Henderson, a subspecies restricted to the Upper Maastrichtian^{6,7}. Since 1965, most biostratigraphers working in the Bay of Biscay region have followed Herm⁸, who used the base of the *Abathomphalus mayaroensis* zone as the base of the Upper Maastrichtian. Thus, the problems perceived in our placement of both the first appearance of *A. mayaroensis* and the Lower/Upper Maastrichtian boundary at 108 m are not real.

In calculating the slope of the strontium evolution curve, we clearly stated our assumptions in going from biostratigraphy to time so that readers could evaluate our reasoning. We used the age determination in ref. 9 as the basis for our absolute age estimate for the duration of the *A. mayaroensis* zone, realizing that other estimates are shorter. Use of a shorter *A. mayaroensis* zone, as postulated by McArthur *et al.*, increases the calculated rate of change through the Upper Maastrichtian and emphasizes the change in slope at around 100 m. This is exactly the opposite to the slower rates of change argued by McArthur *et al.*

McArthur *et al.* argue that fault repetition of the section can explain the isotope trends. The break they identify at 30 m in their Fig. 1 is the same as that we showed in our stratigraphic section. They claim the overlap is 40 m, but base their conclusion, not on any stratigraphic or palaeontological data, but on their preconceived notion that Sr isotope evolution of sea water before the K/T boundary must be monotonic. Their own nannofossil data³ permit a maximum overlap of only 30 m. They also ignored an independent study¹⁰ based on samples from four DSDP cores which found the same pattern of non-monotonic isotope change just below and at the K/T boundary which we presented. Furthermore, we have previously demonstrated that the purported overlap suggested by McArthur *et al.* is false, since *Anapachydiscus terminus* n. sp. of refs 5,7 (= *Anapachydiscus fresvillensis* of ref. 12) occurs through 40 m of the partial section containing the K/T boundary, but only in the top 10 m of the thicker partial section shown in our Fig. 1 (ref. 2). Lithological evidence also shows that the purported stratigraphic overlap has not occurred because the conspicuous

soft brown marl for the upper 11 m of section cannot be found in the thicker of the partial sections shown in our Fig. 1. The section is indeed cut by numerous faults, but these are minor, and can be easily resolved in the field.

McArthur *et al.* implied significant repetition in the lower part of the section. This repetition, however, is impossible because of differences in fossil content (ammonites, inoceramid bivalves, planktonic foraminifera and nannofossils) and lithology between intervals they suggest are repeated. There are five distinctive lithostratigraphic units in the Maastrichtian strata in the Bay of Biscay regions^{5,7,13}, which makes it easy to locate oneself. Our stratigraphic sections of the Bidart locality have been examined independently, and our interpretation verified, by J. Mount (University of California, Davis) and J. Smit (Free University, Amsterdam). Additionally, W. J. Kennedy (Oxford University) has logged the section with us, and is a co-author of two papers^{3,7} using our stratigraphic interpretation.

The lowest break in the section they proposed is not documented in the study¹¹ cited by McArthur *et al.* The closest fault correlates to 110 m, not 125 m, on our section. The placement of this break and the one at 95 m in their figure depends on their assumption of a constant rate of change in the strontium curve and misrepresents our data. Their claim for the lower “isotope discontinuities” is only possible by ignoring analytical uncertainty. The sizes of the two lower discontinuities are less than 50% of our quoted analytical error. If McArthur *et al.* had reproduced our diagram with analytical errors, as originally presented, their insistence on discontinuities would immediately have been seen as untenable.

Peter D. Ward

Bruce K. Nelson

Kenneth G. MacLeod

Department of Geological Sciences,

University of Washington, Seattle,

Washington 98195, USA

1. McArthur, J., Burnett, J. & Hancock, J. *Nature* **355**, 28 (1992).
2. Nelson, B., MacLeod, K. & Ward, P. *Nature* **351**, 644–647 (1991).
3. Burnett, J., Kennedy, W. & Ward, P. *Newslett. i Stratigr.* **26**, 145–155 (1992).
4. Wiedmann, J. *Revista Española de Paleontol. Extraordinario*, 127–140 (1988).
5. Ward, P. & Kennedy, W. *J. paleont. Mem.* (in the press).
6. Henderson, R. & McNamara, K. *Paleontology* **28**, 35–88 (1985).
7. Ward, P., Kennedy, W., MacLeod, K. & Mount, J. *Geology* **19**, 1181–1184 (1991).
8. Herm, D. *Zeit. Deut. Geol. Ges.* **115**, 277–348 (1965).
9. Kent, D. & Gradstein, F. *Bull. geol. Sci.* **96**, 1419–1427 (1985).
10. Martin, E. E. & Macdougall, J. D. *Earth planet. Sci. Lett.* **104**, 166–180 (1991).
11. Clauser, S. C. R. *Acad. Sci., Paris* **304**, Ser. II, 11579–11584 (1987).
12. Ward, P. *Revista Española de Paleontol. Extraordinario*, 119–126 (1988).
13. MacLeod, K. & Ward, P. *Geol. Soc. Am. Sp. Pap.* **247**, 509–518 (1990).

Signac's secret

John D. Mollon

Paul Signac and Color in Neo-Impressionism. By Floyd Ratliff. *Rockefeller University Press*: 1992. Pp. 317. \$49.95.

In the course of his lifetime, Paul Signac owned 32 boats. He chose to live by the River Seine and on the coast of Brittany, and he finally settled in St Tropez. He was fascinated by water. But it was neither the surging ocean that attracted his brush, nor the calm of sheltered lakes; rather it was the dappled, intricately ruffled waters of rivers, canals and ports. And the neo-impressionist technique served well to capture the fragmented, tachistoscopically changing images of objects and lights reflected in such waters. At the hand of Signac, the dots of Seurat's pointillism became the near-rectangular *taches* of divisionism.

But Signac had a grander rationale for

the vibrant mosaic of his paintings. The neo-impressionists claimed to be the first artists to apply scientific principles to painting. Signac, their spokesman, had read a translation of Ogden Rood's *Modern Chromatics* and had once called on Michel-Eugène Chevreul, the centenarian savant of French colour science. By Signac's account, the precepts of neo-impressionism were, first, to avoid the subtractive mixing of pigments on the palette and instead to gain luminosity and lustre by the "optical mixture" of adjacent *taches* of pure colour; second, to use this technique to capture the subtle variation of colour that occurs across the surface of natural objects; and

third, to achieve colour harmony by contrast of complementaries, as advocated by Chevreul.

To this day, art historians are exercised by two distinct questions. How far did Signac really understand, and draw upon, the visual theory of his day? And can the visual science of our own time help us to understand the successes and shortcomings of neo-impressionism? Both questions are tackled in a new book by the distinguished visual scientist Floyd Ratliff. The book is sumptuously illustrated and includes a masterly translation by Willa Silverman of Signac's *D'Eugène Delacroix au néo-Impressionnisme*.

After a short biography of Signac, Ratliff reviews the history of colour science, leading the reader firmly through the conceptual unravelling that has characterized the field. He shows how the trichromatic nature of colour space came slowly to be seen as a property of man rather than as a property of the physical world.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Les Moulins à Overschie (The Windmills at Overschie) by Paul Signac. 1905. Oil on canvas, 25×37¼ inches. Reproduced by permission of the Museum of Fine Arts, Springfield, Massachusetts, USA.

Ratloff ends by offering a 'modern' account of colour theory — a reconciliation of the theory of Sir Thomas Young (*sic*) and that of Ewald Hering. Reconciliations of this kind were fashionable in textbooks 20 years ago and Ratloff shows himself to be out of date. He confounds three types of chromatic antagonism: (1) the phenomenologically opponent pairs of colours, red and green, blue and yellow; (2) complementary pairs of colours, that is, pairs that mix to form white; and (3) the colours that maximally polarize the chromatically opponent channels of the early visual system. In fact (contrary to Ratloff's definitions on page 300), pure red and pure green light mix to form yellow, not white; and the complementary of pure blue is an orange, not yellow. And no one has found cells in the primate visual system that correspond to the red-green and yellow-blue processes of Hering. The two most common types of chromatically

Neo-Impressionists made any deliberate use of the spreading effect, based directly on scientific knowledge of the phenomenon." Ratloff himself puts the emphasis on contrast of adjacent *taches*. I think this is simply wrong. Assimilation is central to Rood's chapter on "The small interval and gradation", which reads as a prescription for neo-impressionism. And I can find no strong basis in Signac's text for Ratloff's claim that Signac confounded optical mixture with contrast. Signac's main use of contrast is between areas larger than the individual *taches*: the contrast of oranges and blues in *Les Moulins à Overschie*, shown on the previous page, offers a captivating illustration.

By adopting *taches* of almost uniform size, the neo-impressionists imposed on themselves one severe limitation: they had less scope to delight the eye with contrast of texture. For lightness and colour are not the only surface properties extracted by the visual system. Perhaps as fundamental is texture, that is, the spatial-frequency content of the stimulus. We use texture, like lightness and colour, to identify objects, and to identify which parts of a scene belong to a common object. And the cells in the visual system that respond to specific spatial frequencies can be interpreted as texture analysers.

Only recently have visual scientists recognized that there is a contrast of texture analogous to the well-known contrast of lightness and colour (S. Klein *et al. Vision Res.* 14, 1421; 1974). There is also 'contrast contrast': a contrast surround will attenuate the perceived contrast of a more delicate texture (C. Chubb *et al. Proc. natn. Acad. Sci. U.S.A.* 86, 9631; 1989). These effects were not explicit in visual textbooks of Signac's day (and are still not understood by galleries that put strongly textured gilt frames around delicately textured compositions). But what the neo-impressionists forwent was certainly understood — either explicitly or implicitly — by those impressionists who eschewed the near-uniform *taches* of Signac. A noble example is offered by Alfred Sisley's *Terrasse à Saint-Germaine: Printemps, 1875*, which is currently hanging in the special exhibition at the Royal Academy in London and in which contrasts of texture interplay with contrasts of hue and lightness.

But why should contrast — of colour, lightness or texture — be so pleasurable to the eye? To this day, visual scientists have no secure answer; and we should be ready to admit it. □

John D. Mollon is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

Chronicles of credulity

Walter Gratzer

On the Wild Side. By Martin Gardner. Prometheus: 1992. Pp. 257. \$24.95.

WHEN the unreason of the irrational inflames a man as much as it does Martin Gardner, one might think that he would do well, for the sake of his health, not to dwell on it too much. But one would be wrong, for like T. H. Huxley, who famously confessed that the flow of bile occasioned by a confrontation with an adversary set him up for days, Gardner thrives on it and transmutes it into often quite wild entertainment. This, his latest collection of essays, is culled partly from his journal, *The Skeptical Inquirer*, and partly from book reviews, some from *The New York Review of Books*, some from less immediately accessible organs such as the *Times-News* of Hendersonville, North Carolina. In them he frequently revisits old friends, first encountered in his classic *Fads and Fallacies* — W. Reich and his orgone box, the hapless parapsychologist J. B. Rhine and the rest of the *galère*. Nor do the tales fade with the telling; rather they ferment and ripen like a fine Stilton.

But the passage of 40 years and the advance of higher education has not abated the tide of human folly. The sage (Nietzsche perhaps?) who opined that conviction is a greater enemy of truth than lies said a mouthful. Here comes the parade of cranks, mediums and occultists — from Aleister Crowley, the Great Beast 666, to the venal televangelists, the Bakkers, Jimmy Swaggart and Oral Roberts and his equally egregious offspring, Richard. Oral, whose Bible fell open before his eyes to reveal the telling passage "Beloved, I wish above all things that thou mayest prosper", eventually misjudged, if not the credulity, then certainly the generosity of his flock, when he announced that if \$14 million were not forthcoming for his latest project, the Lord would "call him home". A group of journalists formed, as I recall, a society called LORD (Let Oral Roberts Die); the faithful failed to fork out and Oral's authority wilted. The doctors at his medical centre did not come up with a promised cure for cancer and the unwanted medical school and hospital are now defunct.

The featherbrained Hopalong and Nancy set the intellectual tone from the White House. The influence of the court astrologers on US affairs of state is now well attested, although the Reagans were clearly embarrassed by the reve-

Reproduced from Paul Signac and Color...



Portrait of Paul Signac by J. Victor (c. 1910).

opponent cell in the early visual system are polarized by red and by blue light and by violet and by yellow light.

Ratloff devotes much discussion to the recognition of spatial contrast by the visual system, but curiously does not emphasize that the analysis is done on different spatial scales simultaneously. The array of retinal photoreceptors is, in fact, examined in parallel by post-receptoral channels tuned to different spatial frequencies: some subsets of cells integrate the input over relatively large retinal areas and are not sensitive to rapid variations in luminance across space, whereas other cells have the task of comparing the local image intensity at adjacent points and are insensitive to slow variation across space.

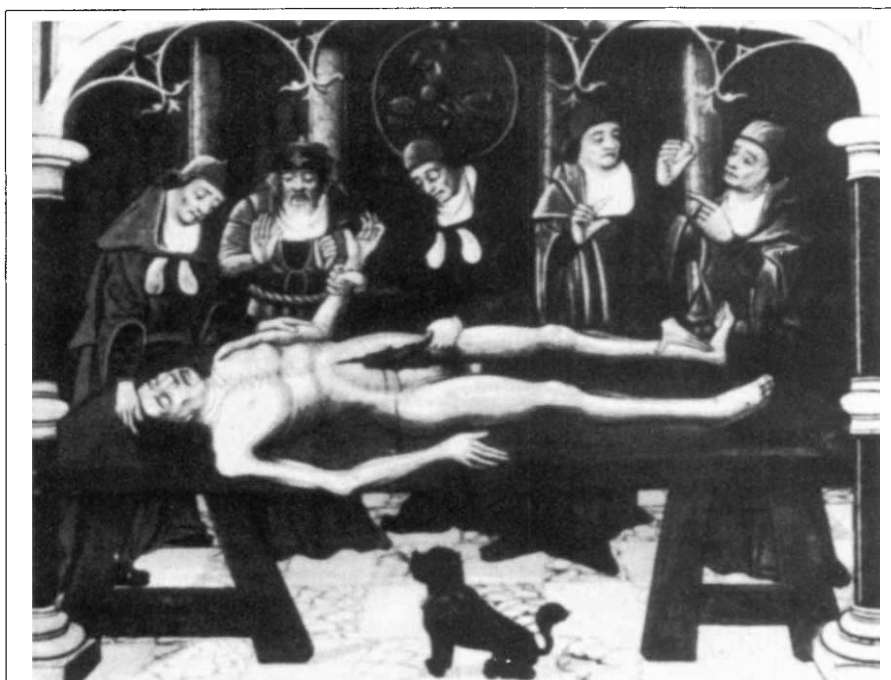
So our visual system can simultaneously show us fine detail, while averaging, say, hue or lightness over a larger area. (For blue and violet colours, chromatic aberration adds to the effect.) This 'spreading effect' or 'assimilation' is central to neo-impressionism. Signac knew that the eye would pick up the vibrant detail of his mosaic while it concurrently averaged colour over several *taches*. Ratloff writes, "It seems unlikely that the

lations and sought to suppress them. A letter in *The New York Times* declared that "the news that important decisions in the White House were based on astrology is most disturbing. The results could undermine faith in astrology." To do the tale full justice would need an H. L. Mencken, but Gardner's account is good enough to make the jaw sag and reason totter on her throne.

One of Gardner's best in a more serious vein is the moral tale of St George Mivart, a heroic Victorian rationalist and a zoologist of high repute, who was also a Roman Catholic. Mivart accepted in essence the Darwinian theory, but perceived that it would give rise to new questions and he prefigured in his writings the operation of a "punctuated" evolutionary mechanism. Mivart strove to persuade the Church to accommodate its doctrine to the new biology, and was eventually excommunicated for his pains; his books were placed on the *Index*, and at the same time spurned by the leaders of scientific opinion for dabbling in religion. Mivart foresaw with clarity the bitter consequences to the Church of its purblind rejection of reality.

Gardner counts James Lovelock among the fantasists and he is none too gentle with J. B. S. Haldane for the evasions over Lysenko that his political commitment forced on him. This is perhaps a touch ungenerous, for Haldane's prevarication stemmed from his loyalty to his friends in the Communist Party, and his discomfort at what was going on in Russia is plain to see from his published writings at the time. Here also we find M. Jacques Benveniste and his homeopathic chemistry of water, and many another comedy. Not all the mildewed cranks, obsessives, zealots and downright lunatics whom Gardner has disinterred point much of a moral, but the results of his researches into the rise of the nineteenth-century mediums are absorbing — especially his meticulous account of the career of the notorious Mrs Piper, who bamboozled William James, but not his more astute sister, Alice. Even the successful practical jokes that Alice James and others played on the mediums and soothsayers did little to shake the faith of the believers. David Hume urged that one should always hold it more likely one had been deceived than that the laws of nature should stand suspended. A quarter of a millennium on — you will find the circumstances in Gardner's book — a reputable British university appoints the first incumbent to the Koestler Chair of Parapsychology. What price progress, eh? □

Walter Gratzer is in the MRC Cell Biophysics Unit, King's College, 26–29 Drury Lane, London WC2B 5RL, UK.



Human dissection (late fifteenth century). Picture reproduced from *The Beginnings of Western Science* by David C. Lindberg, a review of which will appear in a future issue of *Nature*. University of Chicago Press, \$57, £45 (hbk); \$19.95, £15.95 (pbk).

Uniquely human

Owen Gingerich

Space, Time and Man: A Prehistorian's View. By Grahame Clark. Cambridge University Press: 1992. Pp. 165. £24.95, \$39.95.

PROFESSOR Clark, a distinguished prehistorian and sometime master of Peterhouse College in Cambridge, here argues that men have, by extending their perceptions of space and time, more than in any other way established their unique identity as human beings. And modern men have sought increasingly to probe the future by extrapolating from the past. Hence a knowledge of the past will better equip people to identify the social forces they wish to control.

The thesis, once stated, seems so simple as to be peddling the obvious. One hopes that in the embellishment there will be something more profound, perhaps the seeds for a few 'Eureka!' insights. But the treatment is at once far-flung and swift. It is fascinating to speculate on the boats that pre-literate seamen *could* have built to abolish the spatial barriers of water, or to notice that early burials give evidence that humankind was looking to the future. So cursory is the treatment, however, that in eight pages we are carried from the Inca road system to the establishment of the BBC television service. There are beguiling snippets of research, giving a fleeting impression of depth, maybe of

art that conceals art, yet the arena is so vast that much of the exercise seems a breathless romp.

I winced in pain on finding the picture towards the end of "Newton's first reflector telescope built in 1688" and another of "Edwin Hubble with 40-inch Schmidt telescope at a time when the recession of the galaxies was accurately measured (1929)". The Royal Society's little reflector (shown in the illustration) is now a bit suspect, as both the wood and the screw threads prove to be anachronistic; but be that as it may, Newton's telescope was "applauded" by the Royal Society members as early as 1672, and the Schmidt telescope had not been invented in 1929, nor was the 40-inch telescope completed until after the Second World War.

In reflecting on my disappointment with the book, two other volumes come to mind. One, *The Discovery of Time* by Stephen Toulmin and June Goodfield, is honoured in the preface. The other is Alexandre Koyré's *From the Closed World to the Infinite Universe*. The first discusses in some depth mankind's increasing perceptions of time, the second the increasing perceptions of space. Each is longer than Clark's, and neither covers the scope of his canvas, but in their more detailed historical analysis each is ultimately more satisfying.

Nonetheless, Clark's unique view is not without reward. His deep appreciation of our historical roots, seen in a very large frame, should give pause to a modernity that tends to ignore history. In this context his discussion of the

introduction of the theory of evolution is particularly thoughtful. The evolutionists of the 1860s "failed to recognize that human societies, by the mere fact of their being human and subject to history, needed to be treated differently from animal species". The statement is worth pondering within the structure of his essay, although unfortunately he scarcely elaborates on any particulars. But he does return to Darwin in the penultimate paragraph, to remind us that Darwin lost his religious faith at the same time as his pleasure in music and pictures and his capacity even to endure the reading of poetry. Although Clark is never explicit, there is an implication that to lose these capacities is to be less than fully human. □

Owen Gingerich is in the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Under the weather

A. S. Goudie

Geomorphic Responses to Climate Change. By William B. Bull. Oxford University Press: 1992. Pp. 326. £45, \$59.95.

In the literature of geomorphology there are some excellent syntheses, some outstanding texts, but only a few classics based on individual research. Some obvious examples of research classics spring to mind — R. A. Bagnold's *Physics of Blown Sand and Desert Dunes* (Chapman and Hall, 1941) is a timeless and durable example of the genre. Such volumes must stand the test of time, be original and scholarly, and have a significance beyond their own inevitably specialized focus. William Bull's book may well come to join their ranks. Although largely based on work in four geographical locations, its methods and results will open up broader horizons to other workers. It also has the stamp of individuality and personal endeavour.

The book is a departure from other books on climatic influences in geomorphology. Bull does not, for example, pursue the effects of long-term climatic change (that is, during the Tertiary period) in the essentially qualitative way espoused by the German school, nor does he seek to catalogue and classify the nature and origin of the 'bizarre' departures from the humid temperate norm in the way that many French and Anglo-American treatments have tended to. Common terms in the book are

'geomorphic thresholds', 'feedback mechanisms' and 'response times to perturbations'. Bull examines the effects of climatic change on fluvial systems over the past 20,000 to 40,000 years in areas that have a range of current mean annual precipitation from 30 to more than 2,000 millimetres. These areas are in Egypt, Israel, the western United States and New Zealand. Some are monolithologic, others multilithologic.

There are some themes and subjects that recur: the importance of weathering and soil formation as a source and store of debris in systems; the variability of varnish and weathering rind formation under different climatic conditions; the processes by which streams depart from equilibrium and initiate processes that lead to the creation of stream terraces; the function of bedload size and quantity in favouring aggradation; changes in basin plant communities in response to climatic change and their impact on sediment mobility; differences in the timing and number of geomorphic events under different climatic regimes; and the progressive exposure of bedrock outcrops in hot deserts as a powerful self-enhancing feedback mechanism that makes some hillslopes sensitive to climatic perturbations.

The book helps us to understand the past. It may also help us to understand the future, for, as Bull himself writes, "Lessons from histories of landscape change provide insight into future impacts of natural or human-induced climatic change on storm patterns, flood frequency, landslides, stability of valley floors, and agricultural productivity". □

A. S. Goudie is in the School of Geography, University of Oxford, Mansfield Road, Oxford OX1 3TB, UK.

Magnetograph

Michael Fuller

Paleomagnetism: Magnetic Domains to Geologic Terranes. By Robert F. Butler. Blackwell Scientific: 1992. Pp. 319. £29.50, \$44.95.

ROBERT Butler's aim of showing how palaeomagnetism works is certainly fulfilled by this excellent book, which provides an introduction to the fundamental principles of the subject with plenty of examples of practical applications to tectonics and geochronology.

As with all authors of books on palaeomagnetism, Butler faces the difficulty of deciding how much background to include on the geomagnetic field and rock magnetism. Some basic magnetism

and the geomagnetic field are covered, and there is a very useful appendix on units in magnetism. An admirable balance is struck between keeping the discussion short enough to be practical and covering rock magnetism in some detail. The physics and chemistry of rock magnetism are complicated and to describe them well is not easy.

Butler discusses the various magnetic energy terms, domain states and magnetization processes, and provides a particularly skilful exposition of Néel theory. I would like to have seen some recognition of the importance of domain nucleation in the magnetization of fine particles, but perhaps that is precisely the sort of detail one cannot cover in an introductory work. Nevertheless, Butler covers the important aspects clearly, providing a sound basis for the later discussion of palaeomagnetism.

The section on the nuts and bolts of palaeomagnetic technique is first class. This is the most practical discussion of palaeomagnetism I have seen. Like A. Cox and B.-R. Hart's *Plate Tectonics, How it Works* (Blackwell Scientific, 1986), it succeeds by taking the reader through the procedures in detail. Butler is well versed in the techniques of measurements and analysis and gives an even coverage of both areas.

The rest of the book covers special topics in rock magnetism and applications of palaeomagnetism. The examples described are all of considerable interest and give a good impression of the contribution of palaeomagnetism to the earth sciences. But I missed any mention of the applications of palaeomagnetism to establishing secular variation and transition fields during reversals, in which palaeomagnetic techniques are pushed to the limit. For this reason alone the applications are of particular interest, but they may also be a key to understanding the geodynamo.

Useful derivations are given in an appendix. Here, as elsewhere, the explicit use of rotation matrices is avoided in favour of the formulae of spherical trigonometry, with all the attendant problems of employing the correct quadrants and signs. This is a pity, as rotation matrices are now familiar to most earth scientists from their use in plate tectonics. A great virtue of this book is that Butler's workmanlike discussion of palaeomagnetism in terms of basic physics and chemistry removes any last vestige of an excuse for treatment of the subject as a black art or form of magic — a regrettable attitude of too many otherwise sensible Earth scientists. □

Michael Fuller is in the Department of Geological Sciences, University of California at Santa Barbara, Santa Barbara, California 93106, USA.

Low-mass stars in the spheroid of our Galaxy

Harvey B. Richer^{*†} & Gregory G. Fahlman[†]

^{*} Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[†] Department of Geophysics and Astronomy, University of British Columbia, Vancouver, Canada V6T 1Z4

The number of stars in the spheroid of our Galaxy appears to increase steeply for smaller masses, with no evidence of a turnover in a simple power-law distribution down to $0.14 M_{\odot}$, the limit of detectability. The measured rotation curve of the Galaxy at the Sun suggests that the mass function cannot be extended beyond $0.05 M_{\odot}$, in which case low-mass stars fall an order of magnitude short of being able to supply the galactic dark matter.

RECENT results¹⁻⁴ indicate that the number of stars as a function of their mass (the mass function) in the globular clusters in our Galaxy rises steeply at low masses. The mass function of one of these clusters (NGC6397) has been determined down to 0.12 solar masses ($M_{\odot}$). This is within a few hundredths of a solar mass of that of brown dwarfs (very-low-mass stars that are incapable of igniting their nuclear fuel). In none of the cluster mass functions thus far measured is there any evidence of a flattening or a turnover of the slope at the lowest masses, so it would seem likely that brown dwarfs exist (or existed) in these clusters. Most theories of formation of the Galaxy have the stellar component of the halo forming together with the globular clusters^{5,6} so these results imply that much of the halo's mass may be present in the form of brown dwarfs and contribute strongly to the galactic dark matter. Direct measurement of the galactic halo mass function for the visible stars would help in confirming this picture, especially if combined with a determination of the radial density profile of low-mass halo objects. As yet we have no estimate of the radial density profile for low-mass metal-poor stars, and it is of interest to establish whether it follows the same functional form as the more massive objects in the visible spheroid (the ' $r^{1/4}$ ' law, where r is distance from the Galactic Centre) or that of the r^{-2} dark-matter distribution.

There have been several recent attempts to measure the mass function of the spheroid directly. Schmidt⁷, assuming a power-law function of the form $N(m) \propto m^{-(1+x)}$ with $N(m)$ the number density of stars per solar mass, and m the mass, found that $x = 2$ was most consistent with his small sample of 18 stars, whereas Chiu⁸ derived a much flatter slope. Richstone *et al.*⁹ suggest that the only way that low-mass stars can contribute significantly to the galactic dark halo is if the mass function turns up sharply below about $0.5 M_{\odot}$, which is exactly what is observed for the globular clusters. We have attempted to derive the mass function and density distribution of stars well removed from the galactic disk with a new data set which is deeper, covers a greater area and has better spatial resolution than that used in ref. 9.

Imaging and analysis

The data consist of a single $2,048 \times 2,048$ pixel CCD field at high galactic latitude ($l = 109^{\circ}$, $b = 73^{\circ}$) covering about 44 arcmin². These images were obtained under superb conditions at the Canada-France-Hawaii Telescope (CFHT) over the nights 7-11 April 1991. Because of the spectral sensitivity of the chip used in the observations (poor blue response), exposures in visible, red and infrared (V, R and I) were secured totalling about 10 h in exposure time distributed roughly equally between the three colours. The final summed image in all the colours has a full width at half maximum (FWHM) for stellar images of $0.7''$. The frames were shifted by up to several tens of arcsecs between exposures so that flat fields could be derived from a median filter of the images in each colour. This technique worked well, producing final frames globally flat to $\sim 0.3\%$. No fringing was apparent on any frames. Twenty photometric standards with a very wide range in $(V-I)$ colour (-0.02 to 4.02) were

observed during the run to calibrate the colours and magnitudes of the program stars. Nonlinear transformation equations were required to transform the magnitudes and colours from the instrumental system to that of the standard one. The analysis was done with the version of DAOPHOT imbedded in IRAF (Interactive Reduction and Analysis Facility). As most of the objects discovered on the frames are galaxies, the fitting techniques contained in DAOPHOT, which use point spread functions, will not return very accurate magnitudes for the galaxies, but here we are concerned with the stars only. In the following discussion we restrict ourselves to objects with $V \leq 25$, $R \leq 24.5$ and $I \leq 24.5$ magnitudes; for brighter objects than these, the data are essentially 100% complete for point sources, so no corrections for incompleteness are required. These limits were established by adding stars of known magnitude into the frames and reducing them again. The recovered fraction of the added stars then provided the completeness statistics.

The selection of a small sample of stars from a huge sample containing mainly galaxies with a low percentage of stars mixed in is the most difficult aspect of this work. At bright magnitudes morphological classification is straightforward and accurate, deteriorating rapidly toward the faintest images. For this reason we used a combination of image shape and colour to select a stellar sample. The procedure was as follows. From the $\sim 5,000$ objects detected as at least a 4σ fluctuation above the sky on the summed image (all colours together), we found 39 objects that are unequivocally stars. These objects passed all the very conservative tests of image shape (FWHM, ellipticity, Kron statistic¹⁰, Jones *et al.* statistic¹¹), selection as a star by FOCAS, visual inspection and having the correct colours for a star. We then used the properties of these 39 stars as a learning sample to help us to search for other potential stellar objects. We found that the simple shape statistic defined by Jones *et al.*, $\log [(I_{\text{peak}} - I_{\text{sky}})/I_{\text{sky}}]$, provided an excellent criterion for selecting compact objects. For the 39 definite stars, this statistic plotted against magnitude is almost dispersionless down to $I = 24.5$ (Fig. 1a). We therefore applied it to the full sample of almost 5,000 detected objects (Fig. 1b), yielding a total of 91 compact objects. Only 49 of these were located within 2σ (that is, within twice their errors) of the colour-colour relation defined for stars in populations I and II. The remaining 42 are presumably mainly compact galactic nuclei or quasars. Compact galactic nuclei or quasars with stellar colours will always be counted as stars in this procedure, and spectroscopy would be needed to exclude them from the sample.

The distances to the objects selected as stars were determined from photometric parallaxes derived by fitting a linear equation for colours and magnitudes to the population II stellar sample defined by Eggen¹²: we obtained the relation $M_V = 4.270 + 3.377(V-I)$ where all colours and magnitudes have been transformed to the Kron-Cousins system¹³. For a given colour, this relation produces an absolute visual magnitude 1.38 magnitudes fainter than the relation for population I stars, and has almost exactly the same slope¹⁴. The recent parallax study of

Monet *et al.*¹⁵ is in rather poor agreement with this calibration, suggesting an absolute magnitude for extreme subdwarfs about 1 magnitude fainter at a given colour than found here. We suspect that their result is appropriate to extremely metal-poor stars, however, and not applicable to the current sample. Our reasoning is as follows. The mean space velocity with respect to the Local Standard of Rest for the Eggen sample is 239 km s^{-1} , whereas for the five stars with measured radial velocities considered by Monet *et al.*, it is 427 km s^{-1} , suggesting a more extreme group. In fact, Monet *et al.* chose their sample specifically to isolate the most metal-poor objects by considering only stars with transverse velocities greater than 200 km s^{-1} (only two have transverse velocities less than 275 km s^{-1}). The objects that we are selecting here are more likely to be near the average metallicity for population II stars. At a given colour, stars poorer in metals are fainter, and this is likely to be the origin of the difference between the Eggen stars and those of Monet *et al.* Further, the calibration derived from the Eggen stars agrees well with the position of the main sequence in the M_V , $(V-I)$ plane for the globular cluster NGC3201, which has intermediate metallicity¹⁶.

The distances to all the 49 stellar objects were then determined from photometric parallaxes by means of the relation mentioned above, irrespective of whether they were thought to be popula-

TABLE 1 Luminosity and mass functions for the galactic spheroid

M_V	$\log \phi$	$\log N(m)$	$\log m$
5-6	$-8.00 (\pm 0.33)$	-6.88	-0.13
6-7	$-7.88 (\pm 0.30)$	-6.78	-0.18
7-8	$-7.31 (\pm 0.23)$	-6.19	-0.23
8-9	$-7.97 (\pm 0.45)$	-6.98	-0.30
9-10	$-7.50 (\pm 0.33)$	-6.61	-0.42
10-11	$-6.82 (\pm 0.22)$	-5.80	-0.58
11-12	$-6.78 (\pm 0.26)$	-5.47	-0.71
12-13	$-6.76 (\pm 0.34)$	-5.23	-0.79
13-14	$-6.00 (\pm 0.34)$	-4.32	-0.87

ϕ is the number of stars $\text{mag}^{-1} \text{pc}^{-3}$, N is the number of stars (solar mass) $^{-1} \text{pc}^{-3}$. The logarithmic errors in $N(m)$ are the same as in ϕ . Units of m are solar masses ($M_\odot$).

tion I, thick disk or population II stars. As we will concentrate here only on those objects that clearly do not belong to the thin disk of the Galaxy, no great error is introduced by this procedure. The nearest star, according to this equation, is at 177 pc from the Sun (correct distance probably closer to 350 pc as it almost certainly is a population I M-dwarf), and the most distant is 59 kpc away. Four extremely red objects ($V-I > 3.8$) were

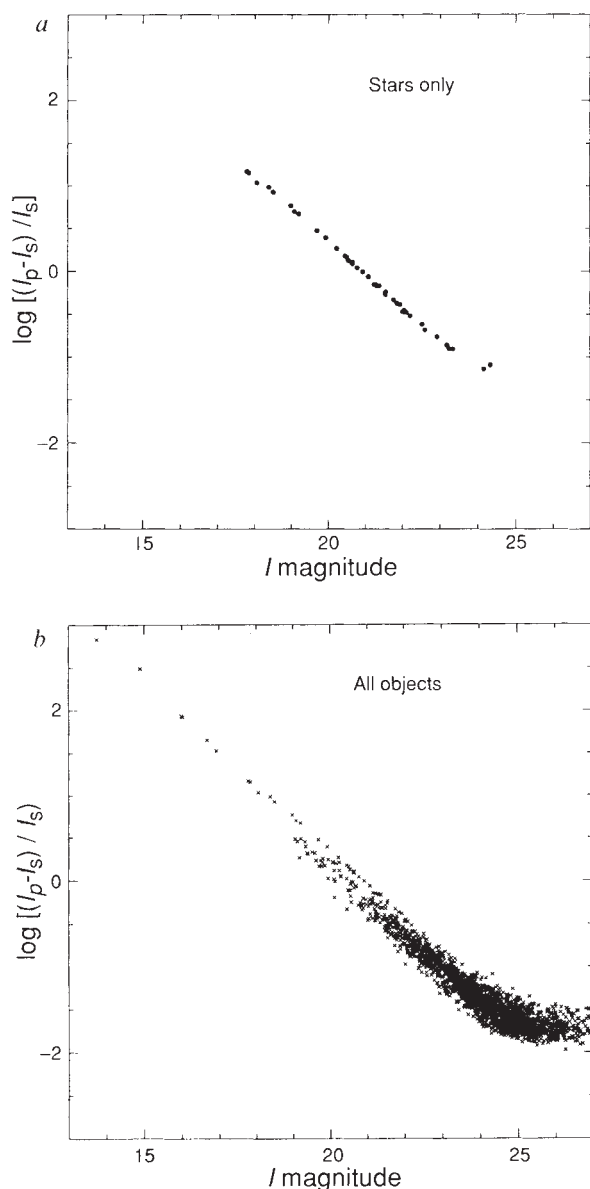


FIG. 1 *a*, Sensitive way to distinguish stars from faint galaxies. The peak intensity in I (minus the sky) divided by the sky intensity is plotted against the I magnitude for each of the 39 objects recognized as stars. Because the intensity distribution of the stars is more peaked than that of galaxies, all galaxies will fall below the relation defined by the stars. *b*, As in *a* but for all the objects to $I=27$ detected in the images. Six slightly saturated stars not included in *a* have been added to the bright end of the distribution. Objects lying below the relation defined by the stars have less peaked intensity distributions and thus presumably are galaxies.

found, but all of these are within ~ 1 kpc of the Sun and are most likely to be low-luminosity population I dwarfs.

Luminosity, mass and density functions

From these data we seek the luminosity function and density distribution of the population II objects. We will convert the luminosity function into a mass function by applying an appropriate mass-luminosity relation for metal-poor stars. In deriving the luminosity function we will consider only those stars which seem to be at distances greater than 1,250 pc from the Sun (corresponding to a height above the plane of $\sim 1,200$ pc). At this height there should be very few of the younger population I stars that make up the thin disk¹⁷, and the stellar population is likely to be dominated by the population II stars belonging to the halo and the spheroid as well as to the thick disk, a proposed component of kinematically older stars that have been distributed to a scale height of about 1 kpc out of the galactic plane by gravitational encounters. This reduces the sample of 49 stars to 31, numbers that are in excellent agreement with the Galaxy model of G. Gilmore (personal communication) which predicts 20 thin-disk objects in our field to $V = 25$ whereas we count 18. In deriving the luminosity (and mass) functions we will not distinguish between the possible thick-disk, spheroid and halo components, but we will assume they are similar. Some justification for this may be found in the mass function for the thick-disk globular cluster M71; although this is currently measured to be flat, there is a suggestion that its present-day slope was produced by dynamical evolution acting on a much steeper initial mass function⁴, more like that seen in the halo globular clusters. Furthermore, the mass function derived for stars more than 5,000 pc from the Sun, which is completely dominated by spheroid and halo objects, is consistent with the one using the larger sample of stars at distances greater than 1,250 pc.

Derivation of the luminosity function and density distribution from a magnitude-limited sample has been extensively discussed^{18–20}. We used the techniques described in ref. 19 and ref. 20, and derived almost identical answers from each method. In Table 1, we list only the results derived by the method of ref. 20, renormalized to agree with those of ref. 19; this is because the former method allows for direct error analysis whereas the latter approach requires Monte Carlo simulations to estimate the errors. The solution for the luminosity function was iterated to an accuracy of 0.1% or better for all ϕ (the number of stars per mag per cubic parsec), and the mass-luminosity relation used to convert the luminosity function to a mass function was that discussed in ref. 1. Assuming a power-law mass function of the form $N(m) \propto m^{-(1+x)}$ for stars with masses less than $0.5 M_{\odot}$, we derive the extremely steep slope of $x = 3.5 \pm 1.2$. By comparison, the Salpeter value for the slope of the mass function in the thin disk of the Galaxy is 1.35. Figure 2a and b shows these luminosity and mass functions; there is no evidence of a flattening or turnover in these functions to the limit of the data, which is near $0.14 M_{\odot}$, within a few hundredths of a solar mass of brown dwarfs for low-metallicity stars. Such a steep mass function would need to extend only a short way into the brown dwarf regime to account for the galactic dark matter. We discuss this point more fully below.

The density function is tabulated in Table 2 and plotted in Fig. 2c. Here we have transformed the distances to those from the Galactic Centre, assuming that the Sun is 8.5 kpc from the Galactic Centre. We have also included in Fig. 2c the density distribution for the galactic spheroid using an $r^{1/4}$ law²¹ (dotted line) scaled as follows. Integrating the disk luminosity function used by Bahcall²¹ to the limit of our data at $M_V = 14$ yields $0.083 \text{ stars pc}^{-3}$ for the local number density of thin-disk stars. We then adopt Bahcall's value of $1/500$ of this for the number density in the spheroid at a radius of 8.5 kpc. Gilmore and Reid²², however, set the spheroid density at $1/800$ of the thin-disk density at this radius, and then include a thick-disk com-

TABLE 2 Density function for the galactic spheroid

$\log \rho$	$(m - M)$	$\log r_e$	$\log R_e$
5.69 (± 0.27)	10.5–11.5	0.23	0.95
5.24 (± 0.24)	11.5–12.5	0.43	0.96
4.97 (± 0.24)	12.5–13.5	0.63	0.99
4.67 (± 0.26)	13.5–14.5	0.83	1.06
4.17 (± 0.28)	14.5–15.5	1.03	1.16
3.69 (± 0.29)	15.5–16.5	1.23	1.30
3.18 (± 0.34)	16.5–17.5	1.43	1.46
2.95 (± 0.37)	17.5–18.5	1.62	1.62
2.70 (± 0.55)	18.5–19.5	1.81	1.81

ρ is the number of stars kpc^{-3} , $(m - M) = 5 \log r/10$ is the distance modulus where r is the distance from the Sun (pc), r_e is the radius from the Sun (kpc) that bisects the volume element and R_e is the radius from the Galactic Centre (kpc) that bisects the volume element.

ponent to make up the same total stellar density. In Fig. 2c there is some evidence that the assumed scaling may be slightly too high in the region of 10–20 kpc, suggesting that if the ratio is taken to be $1/800$ there is room for a thick disk with the properties proposed by Gilmore and Reid. The current data set, however, does not allow us to decide these issues. The main point of Fig. 2c is that the small but faint sample observed here seems to have the expected density distribution, providing confidence in the luminosity function results. The last two points in Fig. 2c do seem to deviate from the $r^{1/4}$ spheroid and suggest a flatter component to the density, closer to the r^{-2} distribution necessary to explain the flat galactic rotation curve. There are, however, only three stars in these two points and the stars are blue, faint and thus liable to misclassification.

Low-mass stars, brown dwarfs and dark matter

We estimate the total mass in the spheroid from the number of stars seen in the volume sampled. To obtain this we use the table published by Young²³ which gives a numerical deprojection of the $r^{1/4}$ law. With a scale radius of 2.7 kpc, we obtain $7.6(\pm 1.9) \times 10^8 M_{\odot}$ which is consistent with the estimates obtained for the standard Bahcall, Schmidt and Soneira (BSS) model^{21,24,25}. This estimate is, of course, only a lower limit to the true mass as we have to add in the unseen contributors: these are remnants (white dwarfs and neutron stars), stars lower in mass than our limit ($0.14 M_{\odot}$) and brown dwarfs. The correction for remnants is uncertain but is unlikely to exceed a factor of 2 (but see Silk²⁶ for an alternative view) whereas that for low-mass stars and brown dwarfs may be substantial.

If brown dwarfs dominate the dark matter content of the Galaxy, estimated to be in the range of $2 - 20 \times 10^{11} M_{\odot}$ (ref. 27), an extension of the observed mass function to $\sim 0.01 M_{\odot}$ is required. Theories of star formation can accommodate such low masses, as the formation of low mass stars down to a minimum of $0.007 M_{\odot}$ has been predicted²⁸ in a metal-poor environment. But the observed Galactic rotation curve provides an important constraint on acceptable mass distributions in the Galaxy, and must be considered in detail.

The two mass models for the Galaxy that we examine here are the four-component BSS model and the three-component Caldwell and Ostriker (C&O) model²⁹ (also discussed in ref. 30). Both models include a disk (the population I component of the Galaxy), a spheroid (the population II component) and a halo (needed to explain the flat outer rotation curve of the Galaxy). The halo is the dominant component in the outer Galaxy and is dark. The fourth component in the BSS model is a compact central mass needed to give the high rotation velocities observed in the inner Galaxy.

These two models differ considerably in detail but both are capable of reproducing the galactic rotation curve. The C&O model predicts that the local spheroid mass density is about 10 times that of the BSS model. Inspection of the rotation curves

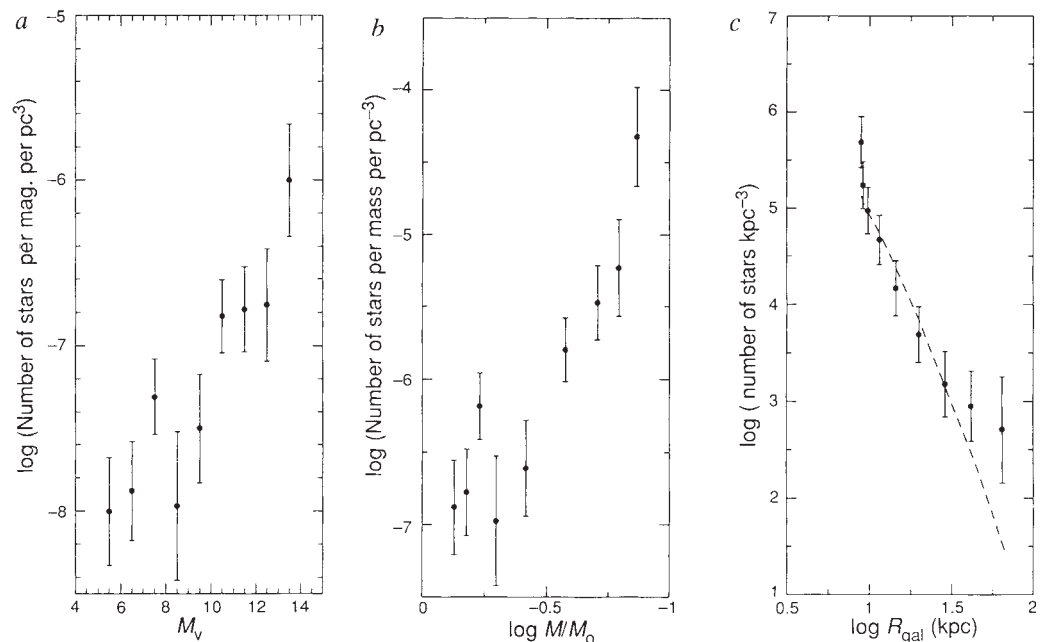


FIG. 2 *a*, Luminosity function of the galactic spheroid. *b*, Derived mass function of the galactic spheroid. *c*, Derived density distribution of stars in the galactic spheroid. The dotted line is the density distribution of the $r^{1/4}$ spheroid scaled as described in the text.

derived from the BSS model^{24,25} suggests that the mass of the BSS spheroid could be increased above the standard value, obtained if one considers the observed stars only, by about a factor of 10 (with a small offsetting decrease in the mass of the BSS central component) without violating the constraint imposed by the galactic rotation curve. To do so would require an extension of our observed spheroid mass function into the brown dwarf regime. For a factor of 10 increase in the observed mass, the corresponding cutoff for our $x=3.5$ mass function would be at $0.056 M_{\odot}$. Our results therefore provide a plausible constituent for the heavy spheroid implied by the C&O dynamical model. Note, however, that our density distribution favours the spheroid density law adopted in the BSS model.

The larger question of whether the extended flat rotation curve of the Galaxy can be accounted for by a population of low-mass stars and brown dwarfs remains unresolved. The dark halo in either the C&O or the BSS model contributes about 10 times as much to the local mass density as does the spheroid (see, for example, ref. 30). Assuming that low-mass stars and brown dwarfs constitute the dark halo, a halo mass function with similar slope and low-mass cutoff to that observed for the spheroid predicts a luminous tail of stars which is not seen near the Sun. Only at distances of more than 40 kpc from the Galactic Centre is there some evidence for an excess of stars above the number expected from the spheroid. At this distance the BSS model predicts that the halo stars should outnumber the spheroid objects by a factor of almost 100; a small deviation (about a factor of 3) is seen in Fig. 2c. At 60 kpc, the observed excess is

about a factor of 10, whereas the excess expected is more than a factor of 100. It may be that the halo mass function is different from that of the spheroid, for example turning up sharply below our observational limit of $0.14 M_{\odot}$ or extending to much lower masses. The alternative explanation may be that low-mass stars and brown dwarfs are not important contributors to the mass budget of the galactic halo.

What seems certain is that the observed stars do follow the $r^{1/4}$ density law assumed for the spheroid in the BSS model, and that the power-law slope of the spheroid mass function is similar to that found in the galactic globular clusters. This latter result is thus strongly suggestive of a similar formation mechanism for the spheroid and for the cluster system, in accord with recent theories of star formation in the early Galaxy^{5,6}. If the observed $r^{1/4}$ density distribution for the spheroid applies to all masses of stars and brown dwarfs, then clearly the mass function cannot be extended to an arbitrarily low limit without violating the constraints imposed by the galactic rotation curve. In the most extreme case of the heavy spheroid in the C&O model, a low-mass cutoff to the spheroid mass function in the region of $0.05 M_{\odot}$ is obtained. A global halo mass function with this low-mass cutoff and similar power-law slope of $x=3.5$ gives a total mass at least 10 times too low to supply the required dark matter content of the Galaxy. It may be that models in which low-mass stellar objects supply the galactic dark matter can only succeed if the stellar luminosity function varies with distance from the centre of the Galaxy, becoming increasingly steep or having a lower mass cutoff at larger distances. □

Received 30 March; accepted 25 June 1992.

1. Drukier, G. A., Fahlman, G. G., Richer, H. B. & VandenBerg, D. A. *Astr. J.* **95**, 1415–1421 (1988).
2. Fahlman, G. G., Richer, H. B., Searle, L. & Thompson, I. B. *Astrophys. J.* **343**, L49–L51 (1989).
3. Richer, H. B., Fahlman, G. G., Buonanno, R. & Fusi Pecci, F. *Astrophys. J.* **359**, L11–L14 (1990).
4. Richer, H. B. *et al. Astrophys. J.* **381**, 147–159 (1991).
5. Lin, D. N. C. & Murray, S. M. in *Formation and Evolution of Star Clusters* (ed. Janes, K.) 55–72 (Astr. Soc. Pacific, San Francisco, 1991).
6. Fall, S. M. & Rees, M. J. *Astrophys. J.* **298**, 18–26 (1985).
7. Schmidt, M. *Astrophys. J.* **202**, 22–29 (1975).
8. Chiu, L.-T. G. *Astr. J.* **85**, 812–823 (1980).
9. Richstone, D., Gould, A., Guhathakurta, P. & Flynn, C. *Astrophys. J.* **388**, 354–361 (1992).
10. Kron, R. G. *Astrophys. J. Suppl.* **43**, 305–325 (1980).
11. Jones, L. R., Fong, R., Shanks, T., Ellis, R. S. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **249**, 481–497 (1991).
12. Eggen, O. J. *Astrophys. J.* **182**, 821–837 (1973).
13. Bessell, M. J. *J. Publ. astr. Soc. Pac.* **91**, 589–607 (1979).
14. Stobie, R. S., Ishida, K. & Peacock, J. A. *Mon. Not. R. astr. Soc.* **238**, 709–727 (1989).
15. Monet, D. G. *et al. Astr. J.* **103**, 638–665 (1992).
16. Brewer, J. thesis, Univ. of British Columbia (1991).
17. Von Hippel, T. & Bothun, G. *Astrophys. J.* (submitted).

18. Lynden-Bell, D. *Mon. Not. R. astr. Soc.* **155**, 95–118 (1971).
19. Choloniewski, J. *Mon. Not. R. astr. Soc.* **226**, 273–280 (1987).
20. Saunders, W. *et al. Mon. Not. R. astr. Soc.* **242**, 318–337 (1990).
21. Bahcall, J. N. A. *Rev. Astr. Astrophys.* **24**, 577–611 (1986).
22. Gilmore, G. & Reid, N. *Mon. Not. R. astr. Soc.* **202**, 1025–1047 (1983).
23. Young, P. J. *Astr. J.* **81**, 807–816 (1976).
24. Bahcall, J. N., Schmidt, M. & Soneira, R. M. *Astrophys. J.* **258**, L23–L27 (1982).
25. Bahcall, J. N., Schmidt, M. & Soneira, R. M. *Astrophys. J.* **265**, 730–747 (1983).
26. Silk, J. *Proc. Nuclear Astrophys. Symp.* (preprint) (1992).
27. Fich, M. & Tremaine, S. A. *Rev. Astr. Astrophys.* **29**, 409–442 (1991).
28. Low, C. & Lynden-Bell, D. *Mon. Not. R. astr. Soc.* **176**, 367–390 (1976).
29. Caldwell, J. A. R. & Ostriker, J. P. *Astrophys. J.* **251**, 61–87 (1981).
30. Kuijken, K. & Gilmore, G. *Mon. Not. R. astr. Soc.* **239**, 571–603 (1989).

ACKNOWLEDGEMENTS. H.B.R. thanks G. Gilmore, B. Boyle, M. Colless, D. Lin, D. Lynden-Bell, R. McMahon, S. White, and T. Von Hippel for valuable discussions, and the Institute of Astronomy, University of Cambridge for hospitality. We thank CFHT for the assignment of telescope time in support of these observations. This research was supported by the Natural Sciences and Engineering Research Council of Canada and by the Killam Foundation. The CFHT is operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique of France and the University of Hawaii.

Induction of germ cell formation by *oskar*

Anne Ephrussi* & Ruth Lehmann

Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Howard Hughes Medical Institute, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

The *oskar* gene directs germ plasm assembly and controls the number of germ cell precursors formed at the posterior pole of the *Drosophila* embryo. Mislocalization of *oskar* RNA to the anterior pole leads to induction of germ cells at the anterior. Of the eight genes necessary for germ cell formation at the posterior, only three, *oskar*, *vasa* and *tudor*, are essential at an ectopic site.

THE ability to reproduce is essential to the survival of any species but, given its prime importance, surprisingly little is known of the molecular and cellular events involved in germ-line formation. Germ plasm and its associated organelles have been described in many organisms¹, but the exact relationship between these structures and the events resulting in the commitment of a group of cells to the germ-cell lineage and maintenance of totipotency is not known. The search for germ-line determinants has not yet indicated which molecules tell cells to take on the germ-cell fate (for a review see ref. 2).

In *Drosophila*, the germ plasm (also referred to as the pole plasm) forms at the posterior pole of the oocyte and contains the determinants for the abdomen and the germ line^{3,4}. The determinant for abdomen formation is the product of the gene *nanos* (*nos*)^{5,6}. One function of the pole plasm is to concentrate *nos* RNA at the posterior pole. Determinants of the germ line, however, have remained elusive. After fertilization of the egg, the first nuclear divisions take place without forming cells. At the tenth nuclear division, nuclei migrate from the centre to the periphery of the egg, where cell formation takes place. The nuclei that migrate into the posterior pole plasm give rise to the posterior pole cells, which form four divisions before the other cells⁷ and are destined to become germ cells. Cytoplasmic transfer of posterior pole plasm to ectopic locations leads to ectopic germ-cell formation, demonstrating that this peculiar cytoplasm possesses germ-cell-inducing potential⁴.

Components of the *Drosophila* germ plasm are synthesized during oogenesis by a cluster of 15 cells, the nurse cells, which are connected to the oocyte at its anterior by cytoplasmic bridges. Pole plasm components are transported into the oocyte and translocated to the posterior pole where they become localized. The abdominal defects associated with lack of pole plasm have allowed the identification of eight maternally active genes, *cap-puccino* (*capu*), *spire* (*spir*), *staufer* (*stau*), *oskar* (*osk*), *vasa* (*vas*), *valois* (*vls*), *mago nashi* (*mago*) and *tudor* (*tud*), whose functions are required for determination of germ-cell fate^{3,8-10,42}.

A number of these genes have been characterized at a molecular level. *Osk* is localized as an RNA to the posterior pole^{11,12}, whereas *stau* and *vas* are localized as proteins^{2,13-15}. We described a pathway for the assembly of posterior pole plasm based on the distribution of *osk*, *Vas* and *Stau* in mutants that disrupt pole plasm formation^{11-13,16,17}. In this pathway, *spir* and *capu* are required for *Stau* localization¹³, *capu*, *spir* and *stau* are necessary for *osk* localization^{11,12}, and *Vas* protein localization depends on *capu*, *spir*, *stau* and *osk* functions^{16,17}. Localization of *Stau*, *osk* and *Vas* is unaffected by *vls* and *tud*, tentatively placing these genes downstream in this pathway. Finally, all of these genes are required for germ-line determination and localization of *nos*, the abdominal signal (L. K. Dickinson, C. Wang and R.L., unpublished results).

Although assembly of these components is stepwise, it is not clear whether all components must be present simultaneously

or whether each component is merely required to recruit the next, so that localization of the last component alone would suffice to trigger germ-line and abdomen formation. To address this, we investigated whether a single component of the pathway is sufficient to trigger assembly of the remaining components. We chose *osk*, because misexpression of *Osk* is easily achieved by mislocalization of its RNA. By mislocalizing *osk* to the anterior of the oocyte, we show that it is sufficient to direct formation of an abdomen and of functional germ cells at an ectopic site. We show that only two genes, *vas* and *tud*, are essential for *osk*-induced pole cell and abdomen formation, whereas *capu*, *spir*, *stau*, and *vls* are not required at the ectopic site. We demonstrate that *osk* dosage determines the final amount of components recruited and ultimately controls the number of pole cells formed.

Number of pole cells is determined by *osk* dose

To investigate the role of *osk* in plasm assembly, we examined the relationship between *osk* gene dosage and amount of pole plasm formed at the posterior of the egg. Specifically, we examined whether an increase from one to four in the number of *osk* gene copies in females would affect the distribution and relative amount of other pole plasm components in their progeny. As expected, a fourfold increase in *osk* gene dosage in females is reflected by a striking increase in *osk* RNA (Fig. 1a, b) and protein (Fig. 1c, d) at the posterior pole of progeny embryos. In addition, a fourfold difference in *osk* copy number results in an increase in the amount of *Vas* protein (Fig. 1e, f), and *Stau* protein (data not shown) at the posterior pole. This relationship between *osk* dosage and the amounts of known pole plasm components at the posterior pole suggests that *osk* is a limiting factor in pole plasm assembly.

If *osk* is limiting for all aspects of pole plasm function, there should be a direct relationship between the amount of *osk* in the embryo and the number of pole cells formed. We counted the number of pole cells in embryos from females with either one or four copies of the wild-type *osk* gene. Embryos derived from females with a single copy of *osk* have on average 10–15 pole cells at cellular blastoderm, whereas those derived from four-copy females have an average of 40–60 pole cells (Fig. 1g, h). This demonstrates that the level of *Osk* protein at the posterior pole determines the number of pole cells formed.

High doses of *osk* cause defects in polarity

Embryos derived from females bearing four copies of *osk* are frequently abnormal. These defects are variable and range from slight malformation of the head to complete lack of head and thoracic structures, and finally, to duplication of abdominal and telson structures at the anterior, the bicaudal phenotype¹⁸ (Fig. 2a–c).

To determine why an increase in pole plasm at the posterior of the egg leads to segmentation defects in the anterior of the embryo, we examined the distribution of three gene products whose activities control segmentation of the embryo: the abdominal determinant *nos*⁵, and the anterior morphogens

* Present address: European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany.

hunchback (*hb*) and *bicoid* (*bcd*)^{18–20}. Because *osk* directs *nos* RNA localization at the posterior pole, we first compared the amount and distribution of *nos* RNA and protein in embryos from females with one or four copies of *osk*. At the higher *osk* dose, a greater quantity of *nos* RNA is localized at the posterior pole (Fig. 2*d, e*). *Nos* protein is also more abundant at the posterior pole and spreads more anteriorly in embryos from four-copy females than from one-copy females (Fig. 2*f, g*).

Because *nos* negatively regulates the translation both of *bcd*

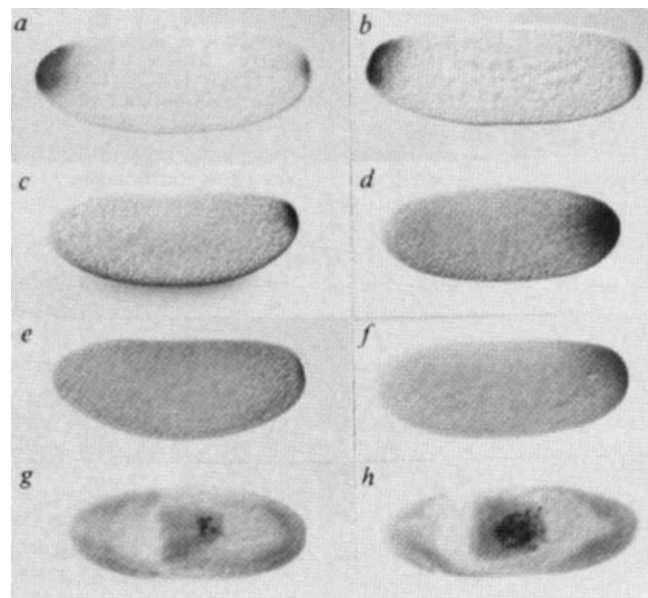


FIG. 1 *Osk* gene dosage determines the amount of pole plasm and pole cells formed. Throughout the text, we refer to embryos from mutant or transgenic females as mutant or transgenic embryos, irrespective of their actual genotype, because all the genes we discuss are maternally active genes. Embryos from females with either one (*a, c, e, g*) or four (*b, d, f, h*) copies of the wild-type *osk* gene. *a, b*, Comparison of the amount and distribution of *osk* RNA in the posterior pole plasm of one and four copy embryos, relative to RNA of the anterior morphogen *bicoid* (*bcd*), which remains constant. *c, d*, *Osk* protein in the pole plasm of one and four copy embryos. *e, f*, *Vas* protein in the posterior pole plasm of one and four copy embryos. *g, h*, *Nos* protein, which is first in the pole plasm of the embryo, is taken up into the pole cells. Pole cells are shown in the midgut pocket, before their migration into the embryo at the extended germband stage. All embryos are oriented anterior left and dorsal up.

METHODS. Single-copy embryos are from *osk*⁵⁴/*TM3* (*a, c, g*) and *Df* (*3R*)^{p^{XT103}}/*TM3* (*e*) females. For the four-copy embryos, a 6.45 kilobase (kb) *XhoI*–*Apal* fragment of *osk* genomic DNA was cloned into the *rosy*⁺ transformation vector, pDM30³⁷. This fragment contains the *osk* gene, including 2.3 kb of upstream and 1.1 kb of downstream sequences. Transformant lines with the P element on chromosome 2 were obtained and made homozygous for the transgene, such that they contain a total of four copies of the wild-type *osk* gene. Two such lines were examined, as well as a line generated in similar fashion but including a 12.75 kb *osk* fragment¹¹. All three lines produce embryos with a variety of head defects. One of the three produces bicaudal embryos. *osk* and *bcd* RNAs were detected by *in situ* hybridization with non-radioactive RNA probes³⁸. Anti-sense probes were prepared by *in vitro* transcription of *osk* and *bcd* complementary DNAs from the T7 polymerase promoters in the *osk*¹¹ and *bcd* c53.46.6³⁹ plasmids. The *Osk* antiserum was produced by rats injected with *Escherichia coli* produced fusion protein consisting of the first 11 amino acids of the T7 phage gene 10 protein and a 294 amino-acid fragment of *osk* (residues 288–582). This portion of the *osk* gene was cloned into the vector pET3c⁴⁰ as a blunted *Styl* fragment. The *Vas* and *Nos* rabbit antisera were gifts of A. Williamson (A. Williamson and R.L., unpublished results) and C. Wang (C. Wang and R.L., unpublished results), respectively. For pole cell counts, one- and four-copy embryos were stained with *Nos* or *Vas*, and the pole cells of 10 embryos of each maternal genotype were counted. Nomarski optics, antibody staining as described⁴¹.

and of *hb*²¹, changes in *Nos* protein distribution could affect the distributions of *Hb* and *Bcd* proteins^{22,23}. Indeed, whereas the *Hb* protein gradient extends into the posterior third of embryos from females with one copy of *osk*, *Hb* protein is restricted to the anterior of embryos from four-copy females (Fig. 2*h, i*). Similarly, whereas *Bcd* protein extends through the anterior third of embryos from females with one copy of *osk*, it is confined to the anterior tip of embryos from females with four copies of *osk* (Fig. 2*j, k*). These changes in the distribution of maternal morphogens account for the defects in patterning of the embryos from females with four copies of the *osk* gene.

As a consequence of reduced *Bcd* and *Hb* activities, genes required for head and thorax formation, which are normally expressed in the anterior half of the embryo^{24,25}, are insufficiently expressed, whereas genes normally expressed in the prospective abdominal region^{26–28} are expressed more anteriorly (Fig. 2*l, m*). This result suggests that increasing *osk* dosage increases the amount of localized *nos* RNA, and that *Nos* protein emanating from the posterior pole can have a far-reaching effect on the entire anterior–posterior pattern.

Pole plasm is assembled by *osk*

Because the amount of *osk* RNA and protein at the posterior pole determines the amount of *Nos* and the number of pole cells, we investigated whether *osk* alone could trigger pole plasm assembly. To provide an ectopic site for potential pole plasm assembly, we mislocalized *osk* RNA to the anterior pole. First, we determined that the 3' untranslated region (3'UTR) of *osk* messenger RNA contains sequences necessary and sufficient for *osk* RNA localization to the posterior of the oocyte (data not shown). We then replaced the *osk* 3'UTR with the *bcd* 3'UTR, which contains the *bcd* anterior localization signal²⁹ (Fig. 3*a*). This *osk-bcd3'UTR* hybrid gene was introduced into the germ line of flies by P element-mediated gene transfer.

The RNA of *osk* is concentrated at both poles of embryos from females bearing the *osk-bcd3'UTR* construct. This bipolar distribution of *osk* RNA is probably the result of wild-type *osk* RNA localized to the extreme posterior pole and hybrid *osk* RNA, like *bcd* RNA, concentrated at the anterior (Fig. 3*b*). As expected, *Osk* protein is found at both poles of *osk-bcd3'UTR* embryos (Fig. 3*c*).

We examined whether localization of *osk* at the anterior of the egg is sufficient to recruit other pole plasm components. *Vas* protein, as well as *nos* RNA, is enriched at the anterior of *osk-bcd3'UTR* embryos (Fig. 3*d, e*). Thus *osk* can trigger assembly of germ plasm components that are normally only found at the posterior pole.

Abdomen formation is induced by *osk*

The *osk-bcd3'UTR* transgene constitutes a dominant female sterile mutation. Females that carry the *osk-bcd3'UTR* transgene produce bicaudal³⁰ embryos, all of which form two posterior abdomens and telsons in perfect mirror image (Fig. 4*a, b*). Formation of the ectopic abdomen at the expense of head structures can be attributed to mislocalization of *nos* RNA (Fig. 3*e*) and protein to the anterior (data not shown), because *osk-bcd3'UTR* embryos that lack *nos* fail to develop either abdomen and form a normal head and thorax (see below).

Germ cells induced by *osk*

If *osk* organizes the pole plasm, it should not only direct abdomen formation, but also induce germ cell formation. Indeed, early embryos from *osk-bcd3'UTR* females form pole cells at both the anterior and posterior poles (Fig. 4*c, d*). The anterior pole cells resemble the posterior pole cells of normal embryos and they contain *Vas* and *Nos* proteins (Fig. 4*e, f*), like pole cells at the posterior pole^{14,15} (C. Wang and R.L., unpublished results).

To test whether the ectopic pole cells are germ-cell precursors, we transplanted anterior pole cells from *osk-bcd3'UTR* embryos

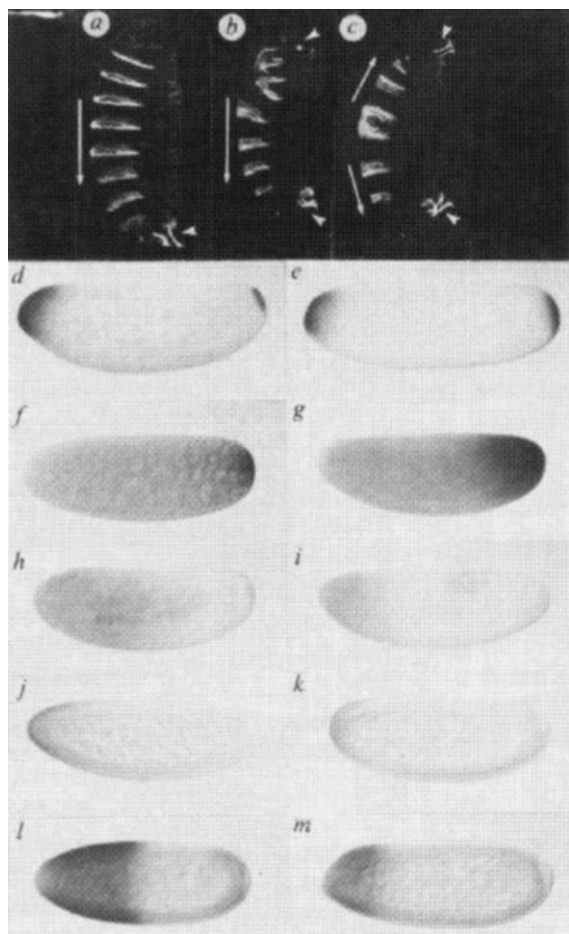


FIG. 2 *Osk* gene dosage determines antero-posterior patterning. *a, b, c*, Cuticles of embryos from females with four copies of *osk*, displaying defects of increasing severity. Of 146 cuticles analysed, 41% had defects in patterning. *a*, Weak phenotype. Embryo lacks most of head and thorax, but has normal abdomen. *b*, Intermediate phenotype. Embryo lacks head and thorax, anterior abdominal segments are irregular, but of normal polarity. Telson is duplicated anteriorly. *c*, Strong phenotype. Bicaudal embryo with two abdomens and telsons in mirror image. Duplication of Filzkörper or Filzkörper material (arrowhead) at the anterior, partial in *b*, and complete in *c*. Anterior to posterior polarity of pattern of larval cuticles is indicated by arrows. Note, embryos from *osk* four-copy females display phenotypes very similar to those of embryos from females mutant for a weak *bcd* allele²⁰. For the wild-type pattern, refer to Fig. 4*a*. Embryos from females bearing either one (*d, f, h, j, l*) or four (*e, g, i, k, m*) copies of the wild-type *osk* gene. *d, e*, Comparison of the amount and distribution of *nos* RNA in the posterior pole plasm, relative to RNA of the anterior morphogen *bcd*, which remains constant³⁹. *f, g*, *Nos* protein in the pole plasm. Note that the *Nos* gradient extends further to the anterior of four-copy than of single-copy *osk* embryos. *h, i*, Hb maternal protein. The Hb gradient is restricted to the anterior of four-copy, as compared with one-copy *osk* embryos. *j, k*, Bcd protein distribution. Bcd protein is either confined to the very tip or not evident in *osk* four-copy embryos. *l, m*, Zygotic Hb protein. Zygotic *hb* transcription is positively regulated by Bcd. Consistent with the reduction in Bcd protein, zygotic Hb protein is confined to the very anterior of *osk* four-copy embryos. The cuticles in *a, b*, and *c* are anterior up and ventral to the left. The embryos in *d–m* are oriented anterior left and dorsal up.

METHODS. *Nos* and *bcd* RNAs were detected by *in situ* hybridization with non-radioactive RNA probes³⁸. Anti-sense probes were prepared by *in vitro* transcription of *nos* and *bcd* cDNAs from the T7 polymerase promoters in the *nos* pN5⁶ and *bcd* c53.46.6³⁹ plasmids. *Nos*, Hb, and Bcd antisera were gifts of C. Wang (C. Wang and R.L., unpublished results), D. Tautz, and W. Driever, respectively. *a, b, c*, Darkfield photographs. *d–m*, Normarski optics; antibody staining as described⁴¹.

into embryos carrying the dominant female sterile mutation *ovo*^{D1}. The *ovo*^{D1} females are sterile and do not produce eggs³¹. We transplanted pole cells from embryos that develop pole cells exclusively at the anterior pole to avoid contamination with posterior pole cells (see legend to Fig. 4). To confirm that the progeny of fertile *ovo*^{D1} females were derived from *osk-bcd3'UTR* anterior pole cells, we observed the development of the eggs of three fertile females. The bicaudal and 'reversed' phenotypes (see below) of the embryos indicated that they could only have resulted from transplantation of *osk-bcd3'UTR*-induced pole cells. Thus pole cells formed at the anterior of embryos from *osk-bcd3'UTR* females can give rise to functional germ cells. We conclude that *osk* induces formation of functional germ cells.

Genetics of pole cell and abdomen formation

The fact that mislocalization of *osk* leads to formation of germ cells and abdomen at the ectopic site (Fig. 5*a, d*) suggests that

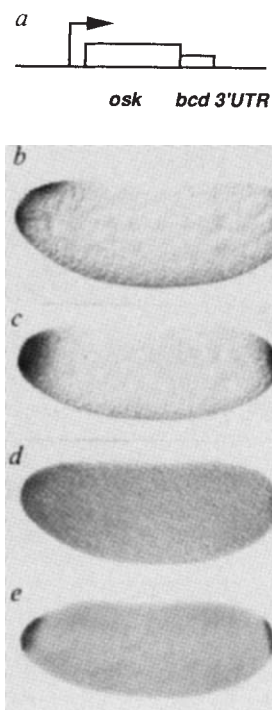
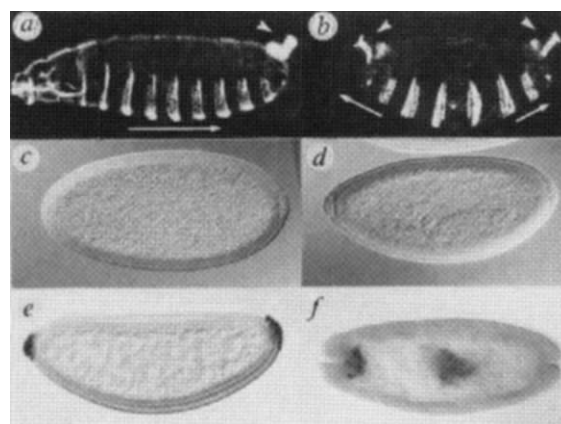


FIG. 3 *Osk* targeted to the anterior pole recruits germ plasm components. *a*, The *osk-bcd3'UTR* hybrid gene. *b, c, d, e*, Distribution of *osk* RNA (*b*), *Osk* protein (*c*), *Vas* protein (*d*), and *nos* RNA (*e*) in embryos from *osk-bcd3'UTR* females. Pole plasm components are localized to the anterior and posterior poles. We also examined the distribution of *Stau* protein and find that its concentration at the anterior is increased in *osk-bcd3'UTR* embryos as compared to wild type. We interpret this enrichment at the anterior as a consequence of a higher concentration of the *bcd3'UTR*, and not as a consequence of *Osk* protein at the anterior (L. Gavis, A.E. and R.L., unpublished results). Embryos are oriented anterior left and dorsal up.

METHODS. The *osk-bcd3'UTR* plasmid was constructed by ligation of a 1.75-kb *EcoRI–BamHI* fragment bearing *osk* promoter sequences from about –1750 to –20 from the transcription start site, a 1.7-kb *BamHI–MluI* *osk* genomic fragment covering the transcription start site to amino acid 448, a 0.5-kb polymerase chain reaction (PCR)-generated *MluI–XbaI* fragment from *osk* amino acid 448 through the translational stop codon TAA, and a 1.7-kb *EcoRI–MluI* fragment containing the *bcd3'UTR*. To avoid possible *nos*-dependent clipping of the fusion transcript, the *nos* response elements^{21,35} are lacking in this *bcd3'UTR*. This construction was transferred into a pDM30³⁷, and injected into embryos from an *osk*⁵⁴ *ry/TM3* line. Previously, we determined that when the *osk3'UTR* is fused to the *lacZ* gene, *lacZ* RNA is localized at the posterior pole of the oocyte (A.E. and R.L., unpublished results). *b–e*, Nomarski optics; RNA *in situ* hybridization detected as in ref. 38.

FIG. 4 *Osk* targeted to the anterior pole induces abdomen and germ-cell formation. **a**, Cuticle of wild-type embryo. The head skeleton is involted and three thoracic segments are followed by eight abdominal segments. Each segment is marked by a denticle belt on the ventral side. The polarity of each segment is indicated by the trapezoidal shape of the denticle belt, with the wider side towards the posterior. The telson at the posterior end carries the light-refractile filzkörper (arrowhead). **b**, Embryo from an *osk-bcd3'UTR* female. All embryos from *osk-bcd3'UTR* females develop symmetrical duplications of the posterior abdomen and telson. The phenotype is very consistent; embryos duplicate the posterior three-and-a-half (**b**) or three (see Fig. 5b) abdominal segments and telson with Filzkörper (arrowhead). **c**, Wild-type embryo at syncytial blastoderm stage. Pole cells are distinct at the posterior pole. **d**, **e**, **f**, Embryos from *osk-bcd3'UTR* females. All embryos have pole cells at the anterior and posterior poles. Embryo in **d**, at syncytial blastoderm. Embryo in **e**, at cellular blastoderm stained for Nos protein. Embryo in **f**, at germband extension stage. Germband extension, which normally only occurs at the posterior pole, is observed at both poles of *osk-bcd3'UTR* embryos. Although the final cuticle pattern is a perfect mirror image, all of the events leading to pole cell formation occur a bit earlier at the posterior than at the anterior pole. At gastrulation, germband elongation is quite asymmetric, with the anterior end lagging the normal posterior. Anterior to posterior polarity of pattern of larval cuticles is indicated by arrows. All embryos are oriented anterior left and dorsal up. **METHODS**. Pole cells were transplanted as described previously³. Pole cell donor embryos were derived from females carrying an *osk-bcd3'UTR* transgene on a chromosome mutant for *osk*⁵⁴, in trans to *osk*⁵⁴ (ref. 3). Sixty-eight



females eclosed and we examined the phenotypes of the progeny of three fertile females. Anterior blastoderm cells of wild-type embryos do not give rise to germ cells⁴. **a**, **b** Darkfield photographs; **c**, **d**, embryos fixed in 4% formaldehyde and photographed with Nomarski optics; **e**, **f**, embryos stained with anti-Nos antibody, as described⁴¹ (antibody was a gift from C. Wang).

genes that are necessary for localizing *osk* RNA at the posterior pole may not be required for germ cell and abdomen formation at the anterior pole. To identify the genes critical for recruitment or activity of germ line and abdominal determinants, we generated females that carried the *osk-bcd3'UTR* transgene and were mutant for *capu*, *spir*, *stau*, *osk*, *vas*, *vls*, *tud* or *nos*. Pole cell and abdomen formation were studied in the progeny of these females. We found that four genes, *capu*, *spir*, *stau*, and *vls* are dispensable for germ cell determination and abdomen formation at the ectopic site. The *osk-bcd3'UTR* females that lack the function of any one of these genes produce progeny of a 'reversed phenotype', with pole cells at the anterior pole and a single abdomen of reversed polarity (data shown for *stau*, Fig. 5b, h; and for *vls*, Fig. 5c, i). Consistent with the role of these genes in posterior pole plasm formation, these embryos lack posterior pole cells and the normal abdomen.

Embryos from *osk-bcd3'UTR* females that lack wild-type *osk* also develop a reversed phenotype (Fig. 5d, j). This demonstrates that the transgene is sufficient to induce pole cells and abdomen. In addition to *osk*, two genes, *vas* and *tud*, are essential for both pole cell and abdomen formation. Embryos from *osk-bcd3'UTR* females that lack *vas* produce embryos that form neither pole cells nor abdomen, and have a normal head, thorax and telson (Fig. 5e, k). *Tud* mutants fail to develop pole cells yet still have residual *nos* activity at both poles, as reflected by the reduced head structures and presence of abdominal segments with normal polarity (data not shown).

Finally, *nos* activity, although essential for abdomen formation, is not required for pole-cell formation. Accordingly, embryos from *osk-bcd3'UTR* females that lack *nos* develop pole cells at both ends and fail to form any abdomen (Fig. 5f, l). In summary, of the genes necessary for posterior pole plasm formation, only three (*osk*, *vas* and *tud*) are essential for germ-cell and abdomen formation at an ectopic site.

Discussion

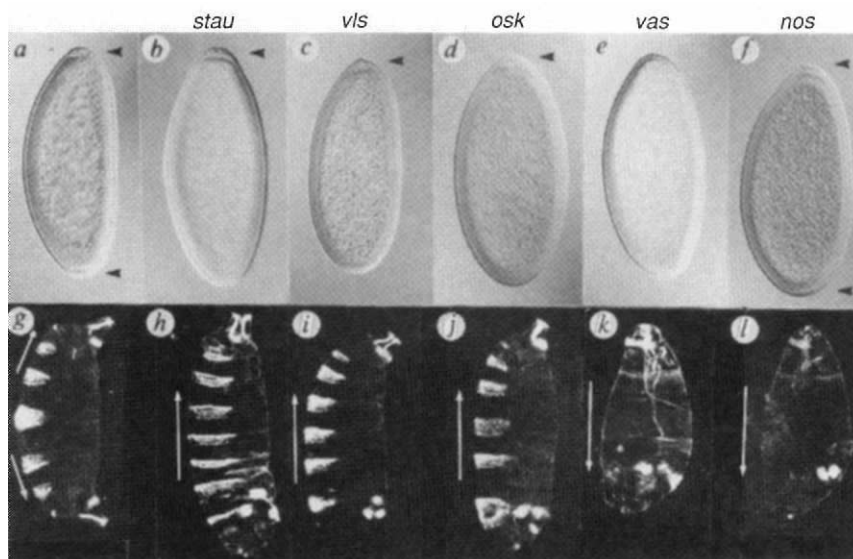
A localized source of *osk* can assemble all components necessary for germ-cell determination and abdomen formation. In addition to *osk*, *vas* and *tud* are essential for germ-cell determination and abdomen formation. The products of the *osk*, *vas* and *tud* genes may participate together in binding *nos* and hypothetical germ-cell determinants; alternatively, because *vas* and *tud* act downstream of *osk*, it is possible that independent localization of one or both of these proteins may bypass a requirement for *osk*.

Capu, *spir*, *stau* and *vls* genes, which are required for posterior pole plasm formation, are dispensable for ectopic germ-cell determination and *nos* localization. Because *capu*, *spir* and *stau* disrupt *osk* RNA localization at the posterior pole, these genes were placed upstream of *osk* in the pole plasm assembly pathway^{11,12}. Our results suggest that the sole function of these genes in pole plasm formation is to transport or tether pole plasm components, like *osk*, to the posterior pole^{5,8}. Because *vls* mutants do not affect *osk* RNA or Vas protein localization, *vls* was placed downstream of *osk* and *vas* in the assembly pathway^{11,12,16,17}. Pole cells do form, however, at the anterior pole of embryos from females that carry the transgene and are mutant for *vls*. Assuming a hierarchical process for pole plasm assembly, it is surprising that a factor that is downstream of *osk* can be dispensable for germ-cell and abdomen formation. Our results suggest that *vls* has a role that is restricted to the posterior pole, and that it is not involved in the early steps of pole plasm assembly (Fig. 6).

There is a direct relationship between the amount of *osk* product and the concentration of pole plasm components, such as Vas and *nos*, at the posterior pole of the embryo. Thus, the increase in concentration of these gene products is independent of their gene dosage. It is unlikely that *osk* controls transcription of these genes, because lack of *osk* does not affect the amount of Vas protein or *nos* RNA in the embryo¹⁷ (C. Wang and R.L., unpublished results). *Osk* gene products could regulate the efficiency of transfer of pole plasm components from the nurse cells to the posterior pole of the oocyte, and such transport might involve association of *osk* products with other pole plasm components in nurse cells. In fact, pole plasm components are assembled into macromolecular structures in the nurse cells^{14,32}. Alternatively, from a uniform pool of pole plasm components in the egg, the dose of *osk* could determine the number of molecules recruited to the posterior pole.

These models do not explain why an increase in *osk* dosage alone causes segmentation defects throughout the embryo. Although we can attribute these defects to elevated levels of *nos* activity, it is unlikely that the total amount of *nos* RNA in the embryo is changed. We favour the idea that the amount of *Osk* protein determines the amount of active Nos protein in the embryo. This could be achieved either selectively, by increasing the amount of Nos synthesized from *nos* RNA localized at the posterior pole or, more generally, by increasing the amount of Nos synthesized from *nos* RNA present throughout the embryo.

FIG. 5 Genetic requirement for pole cell and abdomen formation. All embryos in this figure are from *osk-bcd3'UTR* females that are also mutant for the indicated genes. *a-f*, Syncytial blastoderm embryos. *g-l*, Larval cuticles. *a, g*, Genotype of female: *osk-bcd3'UTR*, otherwise wild-type. Pole cells at anterior and posterior poles (*a*); symmetric bicaudal embryo (*g*). *b, h*, Genotype of female: *osk-bcd3'UTR*; *stau^{D3}/stau^{D3}* (ref. 5). Pole cells at the anterior (*b*); reversed abdomen and duplicated telson at the anterior (*h*). *c, i*, Genotype of female: *osk-bcd3'UTR*; *vls^{PE36}/Df(2R)TW2* (ref. 9). Pole cells at the anterior (*c*); reversed abdomen and duplicated telson at the anterior (*i*). *d, j*, Genotype of female: *osk-bcd3'UTR*; *osk⁵⁴/osk⁵⁴* (ref. 3). Pole cells at the anterior (*d*); reversed abdomen and duplicated telson at the anterior (*j*). *e, k*, Genotype of female: *osk-bcd3'UTR*; *vas^{D1}/vas^{PD}* (refs 5, 9, 15). No pole cells at either pole (*e*); normal head, thorax, and telson, no abdomen (*k*). *f, l*, Genotype of female: *osk-bcd3'UTR*; *nos^{L7}/nos^{L7}* (ref. 5). Pole cells at both poles (*f*); normal head, thorax, and telson, no abdomen (*l*). Note that pole cells at the anterior do not seem to interfere with formation of the head, because the head skeleton of the larvae of this maternal genotype develops normally. Heterozygosity for *nos* in *osk-bcd3'UTR* females results in bicaudal embryos with more abdominal segments duplicated. The dosage of *nos* thus seems to be limiting for abdomen, when *nos* RNA is localized at both poles. Reversed phenotypes were also observed in progeny of *osk-bcd3'UTR*; *capu^{RK}/capu^{RK}* (ref. 8) and *osk-bcd3'UTR*; *spir^{RP}/spir^R* (ref. 8) females. Embryos from *osk-bcd3'UTR*; *tud^{WC}/tud^{WC}* (ref. 10) females form no pole cells and develop into larvae of normal polarity. In these larvae, head structures are reduced and fewer abdominal segments formed. This reflects that *tud* mutants have residual *nos* activity. We also tested the recently described gene *mago nashi* (*mago*)⁴². Embryos from *mago* females fail to form pole plasm and pole cells at the posterior and develop variable abdominal defects. Embryos from *mago¹/mago¹*; *osk-bcd3'UTR* females form pole cells at the anterior and show variable abdominal phenotypes including the 'reverse phenotype'. We therefore conclude that *mago* is not required for pole cell or abdomen formation at the anterior. All embryos are oriented anterior up and ventral left. Pole cells are indicated with an arrowhead. Anterior to posterior polarity



of pattern of larval cuticles is indicated by arrows.

METHODS. *a-f*, Embryos fixed in 4% formaldehyde and photographed with Nomarski optics. *g-l*, Darkfield photographs. The vitelline membrane whose anterior pole is marked by the micropyle, was removed from all embryos shown. But, to ascertain location of pole cell formation and polarity of cuticle, living embryos were observed from each maternal genotype, and their cuticle examined inside the vitelline membrane. Because the *osk-bcd3'UTR* transgene causes dominant female sterility the *osk-bcd3'UTR* lines were maintained through males and *osk-bcd3'UTR* males were crossed to females of the appropriate genotype to generate the mutant combinations described. For combinations with mutations that map to the second chromosome (*capu*, *spir*, *stau*, *vas*, *vls*, *mago*, *tud*) we used a line with integration of the transgene into the third chromosome, whereas for combinations with mutations that map on the third chromosome (*osk*, *nos*) we used a transgenic line with insertion of the transgene into the second and third chromosome produce embryos with identical phenotypes.

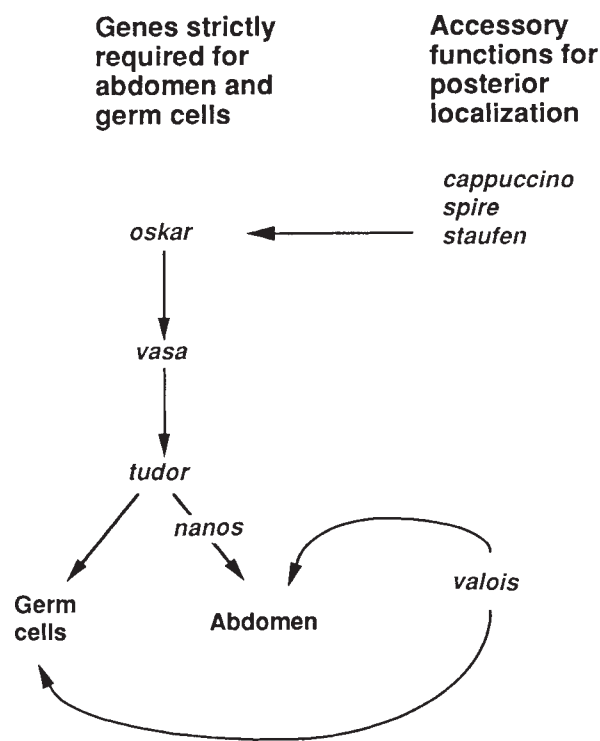


FIG. 6 Pathway for germ-cell and abdomen formation.

Because *osk* concentration at the posterior pole determines the amount of pole plasm and active Nos protein, a further increase in *osk* dose should lead to yet more pole cells and an even more striking effect on pattern formation. Eventually, however, factors such as *Vas*, *nos* and *Tud* must become limiting.

The *osk* RNA is mislocalized in embryos from females that carry dominant gain of function mutations in the *BicaudalD* gene (*BicD^D*)¹¹. Whereas these embryos localize *nos* to both poles¹¹ and develop the bicaudal phenotype^{30,33}, they do not form pole cells at the anterior^{5,34} (A.E. and R.L., unpublished results). It is unlikely that the altered *BicD^D* (refs 35, 36) protein inhibits pole cell formation because embryos from females with both the *osk-bcd3'UTR* transgene and a *BicD^D* mutation form pole cells at the anterior (data not shown). We favour the idea that the amount of anterior *osk* is simply not sufficient for pole cell formation in *BicD^D* embryos. In support of this idea, *Vas* is not concentrated at the anterior pole of *BicD^D* embryos^{16,17}.

By targeting *osk* RNA to an ectopic site, we have determined that only a subset of the genes involved in posterior pole plasm formation is required for germ-cell determination and abdomen formation. Although it is clear that the functional aspects of pole plasm can be duplicated by mislocalization of *osk*, we do not yet know whether the morphological specializations characteristic of the posterior pole plasm, like polar granules, are present at the ectopic site.

Spatially or temporally inappropriate protein expression is often achieved by fusing foreign transcriptional control elements or protein targeting sequences to a gene of interest. In certain circumstances, however, it may be useful to relocate RNAs

within a cell. We have used a localization sequence contained in *bcd* RNA to target *osk* RNA to a novel location. A similar

approach will be useful to identify additional molecules crucial to germ-cell determination. □

Received 15 May; accepted 22 June 1992.

1. Eddy, E. M. *Int. Rev. Cytol.* **43**, 229–280 (1975).
2. Lehmann, R. *Curr. Opin. Genet. Dev.* (in the press).
3. Lehmann, R. & Nüsslein-Volhard, C. *Cell* **47**, 141–152 (1986).
4. Illmensee, K. & Mahowald, A. P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1016–1020 (1974).
5. Lehmann, R. & Nüsslein-Volhard, C. *Development* **112**, 679–691 (1991).
6. Wang, C. & Lehmann, R. *Cell* **66**, 637–648 (1991).
7. Rabinowitz, M. J. *Morph.* **69**, 1–49 (1941).
8. Manseau, L. J. & Schüpbach, T. *Genes Dev.* **3**, 1437–1452 (1989).
9. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302–317 (1986).
10. Boswell, R. E. & Mahowald, A. P. *Cell* **43**, 97–104 (1985).
11. Ephrussi, A., Dickinson, L. K. & Lehmann, R. *Cell* **66**, 37–50 (1991).
12. Kim-Ha, J., Smith, J. L. & Macdonald, P. M. *Cell* **66**, 23–35 (1991).
13. St Johnston, D., Beuchle, D. & Nüsslein-Volhard, C. *Cell* **66**, 51–63 (1991).
14. Hay, B., Ackerman, L., Barbel, S., Jan, L. Y. & Jan, Y. N. *Development* **103**, 625–640 (1988).
15. Lasko, P. F. & Ashburner, M. *Nature* **335**, 611–617 (1988).
16. Lasko, P. F. & Ashburner, M. *Genes Dev.* **4**, 905–921 (1990).
17. Hay, B., Jan, L. H. & Jan, Y. N. *Development* **109**, 425–433 (1990).
18. Struhl, G. *et al. Cell* **69**, 237–249 (1992).
19. Lehmann, R. & Nüsslein-Volhard, C. *Dev. Biol.* **119**, 402–417 (1987).
20. Frohnhofer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120–125 (1986).
21. Wharton, R. P. & Struhl, G. *Cell* **67**, 955–967 (1991).
22. Tautz, D. *Nature* **332**, 281–284 (1988).
23. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 83–93 (1988).
24. Tautz, D. *et al. Nature* **327**, 383–389 (1987).

25. Gaul, U., Seifert, E., Schuh, R. & Jäckle, H. *Cell* **50**, 639–647 (1987).
26. Nauber, U. *et al. Nature* **336**, 489–492 (1988).
27. Eldon, E. D. & Pirrotta, V. *Development* **111**, 367–378 (1991).
28. Kraut, R. & Levine, M. *Development* **111**, 601–609 (1991).
29. Macdonald, P. M. & Struhl, G. *Nature* **336**, 595–598 (1988).
30. Bull, A. J. *exp. Zool.* **161**, 221–241 (1966).
31. Oliver, B., Perrimon, N. & Mahowald, A. P. *Genes Dev.* **1**, 913–923 (1987).
32. Mahowald, A. P. *J. exp. Zool.* **176**, 329–344 (1971).
33. Mohler, J. & Wieschaus, E. F. *Genetics* **112**, 808–822 (1986).
34. Mohler, J. & Wieschaus, E. F. *Cold Spring Harb. Symp. quant. Biol.* **50**, 105–111 (1985).
35. Wharton, R. P. & Struhl, G. *Cell* **59**, 881–892 (1989).
36. Suter, B., Romberg, L. M. & Steward, R. *Genes Dev.* **3**, 1957–1968 (1989).
37. Spradling, A. C. in *Drosophila: a Practical Approach* (ed. Roberts, D. B.) 175–197 (IRL, Oxford, 1986).
38. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
39. Berleth, T. *et al. EMBO J.* **7**, 1749–1756 (1988).
40. Rosenberg, A. H. *et al. Gene* **56**, 125 (1987).
41. Macdonald, P. M. & Struhl, G. *Nature* **324**, 672–675 (1986).
42. Boswell, R. E. *et al. Development* **113**, 373–384 (1991).

ACKNOWLEDGEMENTS. We thank J. Moore for technical assistance, W. Driever, D. Tautz, C. Wang and A. Williamson for gifts of antibodies, L. Dickinson, L. Gavis and C. Wang for sharing their unpublished results, the members of the Lehmann laboratory for discussions, S. Burden, L. Gavis and C. Wang for comments on the manuscript and A. Kron for help with its preparation. A.E. would like to thank H. Stefánsson. This study was supported by the NIH and the Packard Foundation. R.L. is a Howard Hughes Medical Institute investigator.

LETTERS TO NATURE

Seasonal cycle of surface ozone over the western North Atlantic

Samuel J. Oltmans* & Hiram Levy II†

*NOAA/Climate Monitoring and Diagnostics Laboratory, Boulder, Colorado 80303, USA

†NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey 08542, USA

THE possible impact of pollution from North America and Europe on tropospheric ozone throughout the Northern Hemisphere is a major environmental concern^{1–4}. We report here continuous measurements of ozone from Bermuda (32° N, 65° W) and Barbados (13° N, 60° W), which suggest that despite their proximity to the eastern US seaboard, natural processes rather than pollution control surface ozone in these regions. Although springtime daily average ozone concentrations at Bermuda are greater than 70 parts per billion (10⁹) by volume (p.p.b.v.) and hourly values in 1989 sometimes exceeded the Canadian Air Quality limit of 80 p.p.b.v., trajectory analyses indicate that these high levels of ozone are transported from the unpolluted upper troposphere >5 km above the northern United States and Canada⁵. During the summer, when surface ozone concentrations over the eastern United States can exceed 70 p.p.b.v. owing to pollution⁶, typical values at Bermuda are between 15 and 25 p.p.b.v. At Barbados, both the seasonal and diurnal variations in surface ozone are nearly identical to those at Samoa in the tropical South Pacific, where the isolation from anthropogenic sources⁷ and low levels of NO_x (ref. 8) ensure that natural processes control surface ozone^{9–11}.

Tropospheric ozone, which controls the chemical cycling of atmospheric trace gases¹² and has a considerable impact on climate¹³, is supplied naturally by downward transport from the stratosphere¹¹ and is also produced photochemically (depending on the local level of nitrogen oxides (NO_x))¹². Over polluted regions of the continental Northern Hemisphere, tropospheric ozone concentrations frequently rise above the natural background^{1,6}, and levels in Europe may have increased by more than a factor of two in the last 100 years¹⁴. Furthermore, in the summer, ozone-rich air transported from Europe has been observed episodically at Mace Head, Ireland¹⁵ and Izania, Canary Islands¹⁶ in the eastern North Atlantic; and we report here some

instances of ozone pollution transported from North America to Bermuda in the western North Atlantic. But the bulk of our observations from the western North Atlantic, when combined with earlier ozone data from the Bahamas¹⁷ and shipboard data from the eastern North Atlantic^{18,19}, suggest that for most of the year it is transport from the stratosphere and natural photochemical destruction, rather than pollution-driven photochemical production, that controls surface ozone levels over much of the North Atlantic.

As part of the Atmosphere/Ocean Chemistry Experiment (AEROCE), we began ozone measurements from 20-m towers at two sites: Bermuda in October 1988 and Barbados in April 1989. The ultraviolet absorption instruments used to measure ozone²⁰ have been intercompared with a network standard instrument maintained by the NOAA Climate Monitoring and Diagnostics Laboratory in Boulder, Colorado. This instrument, in turn, is compared to the US National Institute for Standards and Technology reference ozone photometer maintained at Gaithersburg, Maryland. Data from American Samoa are tied to the same standard. Measurements are made at 20-s intervals and averaged into hourly values.

The monthly average ozone concentrations at Bermuda (Fig. 1) show a marked seasonal cycle, with median values reaching 50 p.p.b.v. in April and dropping to below 20 p.p.b.v. in August. The large monthly variabilities result from frequent multi-day transport events (see insets *a* and *b* in Fig. 1) and from year-to-year differences. Throughout the winter and spring there are a number of multi-day events when hourly concentrations exceed 70 p.p.b.v., whereas in the summer, peak hourly values seldom reach 50 p.p.b.v. and the daily average dips below 15 p.p.b.v. This deep summer minimum is in marked contrast to surface observations over the eastern United States, where as a result of regional pollution, the monthly means range from 35 to 50 p.p.b.v. and afternoon hourly averages frequently exceed 70 p.p.b.v. (ref. 6). The seasonal cycle observed at Bermuda is qualitatively similar to earlier observations from the Bahamas¹⁷ (also in the western North Atlantic at 21.5° N, 70° W) and to ship cruise data from the eastern North Atlantic at 30–40° N (ref. 18).

To differentiate between anthropogenic and natural control of ozone concentrations, we use isentropic back-trajectory analysis to identify the sources of both the high and the low ozone mixing ratios. These trajectories, which trace back in time the

three-dimensional path that an adiabatic air parcel would have taken to reach the measurement site^{21,22}, are calculated at 12-hour intervals from the National Meteorological Center's global analysis of meteorological data²³. They follow the flow of the large-scale winds in the stably stratified atmosphere and must remain above the non-adiabatic mixed layer (above 950–850 hPa). In this analysis, it is assumed that any parcel descending to the top of the mixed layer over the measurement site will be transported down to the tower at 20 m by boundary-layer turbulence. Although such trajectories are a simplification of the complex meteorology, they can be used to identify the general direction of transport, to differentiate between air that has remained in the lower troposphere and air that has descended from the mid- and upper troposphere, and to estimate the speed at which the tracer has travelled to the measurement site.

All of the high springtime ozone events (>50 p.p.b.) are associated with back-trajectories that start 5 to 10 days earlier north and west of Bermuda, at least a few kilometres above the surface. Back-trajectories associated with the highest ozone mixing ratios originate north of 50° N at 500 mbar (5.5 km) or higher, where in the spring 'tropopause fold' events have been directly observed to carry ozone-rich air down from the lower stratosphere²⁴ and the monthly average mixing ratio exceeds 60 p.p.b.v.^{1,5}. This picture is consistent with ozone measurements from a ship cruise in the eastern North Atlantic. Smit *et al.*¹⁹ found high springtime surface values between 25° and 40° N that seemed to have come from the upper troposphere. During the summer, atmospheric transport to Bermuda is dominated by a large anticyclone (the Bermuda high) and air parcels travel slowly in the lower troposphere over the subtropical and tropical Atlantic. As a result, the ozone levels decline throughout the summer and often dip below 15 p.p.b.v. (see inset *b* in Fig. 1). There are occasional influxes of aged pollution, which have travelled slowly from the eastern United States (see the four peaks between 35 and 45 p.p.b.v. in inset *b*).

A particularly striking transport event occurred during the passage of a cold front on 19 May 1990 (see inset *a* in Fig. 1). Ahead of the front on 18 May, the near-surface wind was from the southwest, the relative humidity was >95% and the ozone concentration was <15 p.p.b.v. The 10-day isentropic back-trajectories associated with these very low ozone levels represent

clean air that has travelled slowly over the ocean south of Bermuda (Fig. 2*b*) and has remained in the lower troposphere below 800 hPa (1.6 km) (Fig. 2*a*). After the passage of the front on 20 May, the wind blew from the northwest, the relative humidity dropped sharply to ~50% and ozone concentration increased rapidly to 70 p.p.b.v. Accompanying the increase in ozone concentration is an equally rapid shift in the isentropic back-trajectories (see Fig. 2*c, d*). They now represent clean air that started in the upper troposphere over western Canada and descended rapidly behind the front as it travelled across the eastern United States and out to Bermuda.

The seasonal pattern at Barbados, which has surface winds blowing almost continuously from the east, differs somewhat from that at Bermuda (Fig. 3). Although both locations have August minima, the Barbados maximum comes in January instead of April, and as can be seen from the boxes and whiskers in Fig. 3, there is much less day-to-day, seasonal and year-to-year variability. The periods of relatively high ozone levels, which are most frequent during the winter and more persistent than

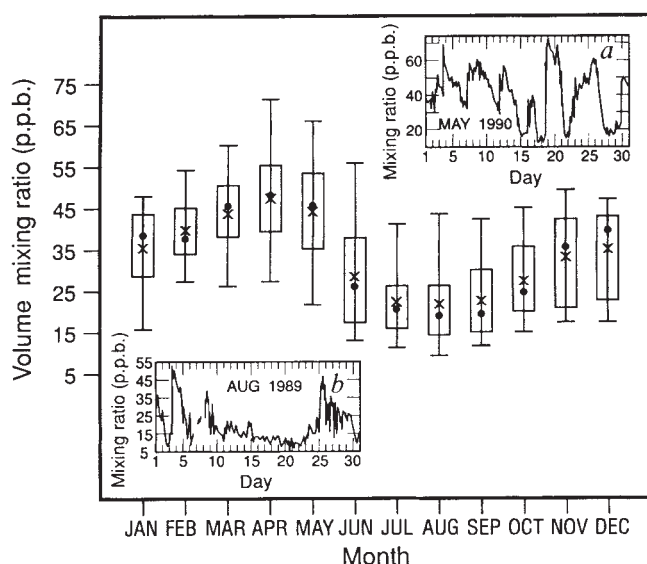


FIG. 1 Average monthly near-surface ozone mixing ratios at Bermuda during the period October 1988 to September 1991. The filled circle is the median, the cross is the mean, the box represents the middle fiftieth percentile and the whiskers the fifth and ninety-fifth percentiles. Inset *a* shows the hourly mixing ratios from May 1990, and inset *b* shows the mixing ratios from August 1989; note the difference in the ordinate scale.

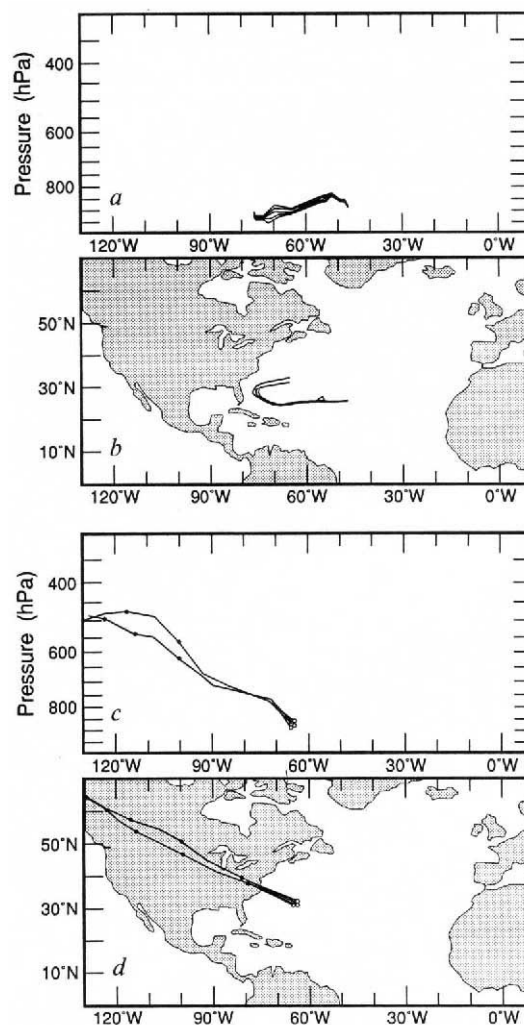


FIG. 2 Ten-day isentropic back-trajectories to Bermuda on 18 May 1990 at 00:00 UT (300 K isentropic surface) are shown in panels *a* and *b*, and those to Bermuda on May 20 1990 at 00:00 UT (298 K isentropic surface) are shown in panels *c* and *d*. Panels *a* and *c* plot the longitude and the altitude of the 10-day isentropic back-trajectories and panels *b* and *d* show the longitude and latitude for the same trajectories. Four trajectories are plotted on each surface, where the appropriate surface is the one just at the top of the atmospheric boundary layer over Bermuda. Details of the trajectory calculation are given in Merrill *et al.*²³. Note that the trajectories for 18 May proceed west of Bermuda to 75° W and then curve cyclonically back to 65° W.

at Bermuda, are associated with atmospheric transport from higher latitudes and altitudes. In the summer and autumn months the ozone levels remain uniformly low, even during the frequent periods of enhanced transport from Africa (D. Savoie, personal communication). Imposed on the relatively low day-to-day variability is a marked diurnal cycle which is easily seen in the hourly time series for March 1990 shown in the inset *b*. This diurnal cycle persists throughout the 3-year record and has an amplitude of ~ 2.5 p.p.b. with a maximum at $\sim 08:00$ local time (LT) and a minimum at $\sim 17:00$ LT (see inset *b*).

Both the diurnal and seasonal patterns at Barbados bear a striking resemblance to those seen at Samoa⁹. The seasonal maxima are roughly twice their respective seasonal minima, and the diurnal amplitudes are 1.8–2.8 p.p.b.v. per day. Oltmans⁹ has already suggested that the diurnal pattern at Samoa, which follows the diurnal change in solar flux, is driven by the net chemical destruction of ozone. Detailed chemical calculations for the tropical South Pacific by Liu *et al.*¹⁰, using the low NO_x levels observed for that region⁸, yield daily ozone net destruction rates of 1.0–2.4 p.p.b.v. per day. This range, which depends on the ozone concentration, agrees with diurnal amplitudes observed at both tropical locations. We infer that Barbados also has extremely low levels of NO_x and a resulting net photochemical destruction of ozone. Although Bermuda's higher synoptic variability (particularly during the large winter and spring transport events) greatly increases the statistical variance, a similar diurnal cycle of ~ 2 p.p.b.v. per day in the hourly means is observed. This, together with the seasonal cycle, suggests that the ozone chemistry at Bermuda is also dominated by photochemical destruction.

Although the diurnal and seasonal patterns are very similar, note that the monthly means at Barbados are 8–10 p.p.b.v. higher than those at Samoa. Two explanations have been offered for this interhemispheric difference, which is also observed over the eastern Atlantic^{18,19} and the Pacific⁵: either there is more effective downward transport of stratospheric ozone in the Northern

Hemisphere^{11,25}, or there is chemical production of tropospheric ozone there, due to a higher background level of NO_x (ref. 26). Although our observations support the dominant role of transport, they may not be representative of the tropospheric column. For example, satellite measurements of tropospheric ozone²⁷ show a summertime maximum over Bermuda, whereas we find a minimum in the atmospheric boundary layer. Indeed, differences in both tropospheric ozone transport^{11,17} and chemistry^{1,28} may cause the seasonal cycle in ozone to vary with altitude. *In situ* systematic measurements of NO_x and ozone vertical profiles to determine the relative contributions of transport and chemistry over the North Atlantic are needed to resolve this question. □

Received 6 January; accepted 5 June 1992.

1. Logan, J. A. *J. geophys. Res.* **90**, 10463–10482 (1985).
2. Liu, S. C. *et al. J. geophys. Res.* **92**, 4191–4207 (1987).
3. Isaksen, I. S. A., Jonson, J. F., Reeves, C. E., Solberg, S. & Chatfield, R. B. in *Ozone in the Atmosphere* (eds Bojkov, R. D. & Fabian, P.) 544–547 (Deepak, Hampton, 1989).
4. Fishman, J., Watson, C., Larsen, J. & Logan, J. *J. geophys. Res.* **95**, 3599–3617 (1990).
5. Oltmans, S. J., Komhyr, W. D., Franchois, P. R. & Matthews, W. A. in *Ozone in the Atmosphere* (eds Bojkov, R. D. & Fabian, P.) 539–543 (Deepak, Hampton, 1989).
6. Logan, J. A. *J. geophys. Res.* **90**, 8511–8532 (1989).
7. Prospero, J. M. & Savoie, D. L. *Nature* **339**, 687–689 (1989).
8. McFarland, M. D., Kley, D., Drummond, J. W., Schmettek, A. L. & Winkler, R. H. *Geophys. Res. Lett.* **6**, 605–608 (1979).
9. Oltmans, S. J. *J. geophys. Res.* **86**, 1174–1180 (1981).
10. Liu, S. C., McFarland, M., Kley, D., Zafiriou, O. & Huebert, B. *J. geophys. Res.* **88**, 1360–1368 (1983).
11. Levy II, H., Mahlman, J. D., Moxim, W. J. & Liu, S. C. *J. geophys. Res.* **90**, 3753–3772 (1985).
12. Crutzen, P. J. in *Tropospheric Ozone* (ed Isaksen, I. S. A. 3–32 (Reidel, New York, 1989).
13. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
14. Volz, A. & Kley, D. *Nature* **332**, 240–242 (1988).
15. Simmonds, P. G. & Derwent, R. G. *Atmos. Environ.* **25a**, 1795–1808 (1991).
16. Schmitt, R., Schreiber, B. & Levin, I. *J. Atmos. Chem.* **7**, 335–351 (1988).
17. Chatfield, R. & Harrison, H. *J. geophys. Res.* **82**, 5969–5976 (1977).
18. Winkler, P. *J. Atmos. Chem.* **7**, 73–91 (1988).
19. Smit, H. G. J., Kley, D., McKeen, S., Volz, A. & Gilge, S. in *Ozone in the Atmosphere* (eds Bojkov, R. D. & Fabian, P.) 419–422 (Deepak, Hampton, 1989).
20. Oltmans, S. J. & Komhyr, W. D. *J. geophys. Res.* **91**, 5229–5236 (1986).
21. Levy II, H. *Nature* **325**, 761–762 (1987).
22. Danielson, E. F. *J. Atmos. Sci.* **25**, 502–518 (1968).
23. Merrill, J. T., Uematsu, M. & Bleck, R. *J. geophys. Res.* **94**, 8584–8598 (1989).
24. Browell, E. V., Danielson, E. F., Ismail, S., Gregory, G. L. & Beck, S. M. *J. geophys. Res.* **92**, 2112–2120 (1987).
25. Mahlman, J. D., Levy II, H. & Moxim, W. J. *J. Atmos. Sci.* **37**, 655–685 (1980).
26. Fishman, J. & Crutzen, P. J. *Nature* **274**, 855–858 (1978).
27. Fishman, J., Fakhruzzaman, K., Cross, B. & Nganga, D. *Science* **252**, 1693–1696 (1991).
28. Pittcock, A. B. *Q. Jl. R. met. Soc.* **103**, 575–584 (1977).

ACKNOWLEDGEMENTS. We thank S. Manabe for helpful comments. The trajectory calculations were performed by J. Merrill and the ozone observations by A. Glasspool and C. Shea.

Cirrus-cloud thermostat for tropical sea surface temperatures tested using satellite data

Rong Fu*, Anthony D. Del Genio†, William B. Rossow† & W. Timothy Liu‡

* Department of Atmospheric Sciences, University of California, Los Angeles, California 90024, USA

† NASA/Goddard Institute for Space Studies, New York, New York 10025, USA

‡ NASA/Jet Propulsion Laboratory, Pasadena, California 91109, USA

RAMANATHAN and Collins¹ have suggested cirrus clouds associated with tropical convection might act as a 'thermostat' to limit tropical sea surface temperatures (SSTs) to less than 305 K by shielding the ocean from sunlight. Here we use satellite radiance data to test this hypothesis. We find that changes in the properties of cirrus clouds do not seem to be related to changes in SSTs. During the 1987 El Niño event, large-scale perturbations to the radiative effects of cirrus clouds were controlled by changes in large-scale atmospheric circulation rather than directly by SSTs. If they are averaged over the entire tropical Pacific, increases in surface evaporative cooling are stronger than decreases in solar

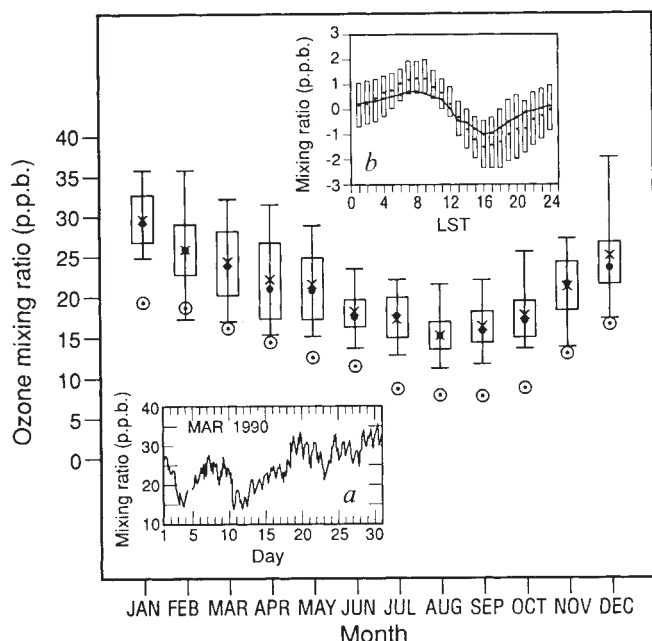


FIG. 3 Average monthly near-surface ozone mixing ratios at Barbados during the period April 1989–September 1991. Symbols are as in Fig. 1. The open circles are monthly median mixing ratios at American Samoa (1973–1991) offset by 6 months. Inset *a* shows the hourly mixing ratios at Barbados for March 1990. Inset *b* shows the average hourly deviations from the daily mean ozone mixing ratio at Barbados (1989–1991). The symbols are as in Fig. 1. The plus signs (+), which are connected by a solid line, are the hourly deviations at Samoa.

heating owing to cirrus cloud variations. Thus we conclude that there is no 'cirrus cloud thermostat' to tropical SSTs.

Ramanathan and Collins¹ expressed their hypothesis as follows: "as the tropical oceans warm, the rapid rise of G_a (the atmospheric greenhouse effect) with SST leads to an unstable feedback. The warming continues until the clouds become thick enough to shield the ocean from the solar radiation and arrest further warming." This hypothesis is based largely on a supposed local correlation between anomalies in cloud forcing obtained from the Earth Radiation Budget Experiment (ERBE) and underlying SST anomalies between El Niño and non-El Niño years. But the tropical atmosphere-ocean system is strongly coupled by dynamics: a perturbation at one location influences other regions of the tropics as well. This fact raises three questions. Is the large-scale behaviour of clouds different from their local behaviour? Does cirrus-cloud variability depend on the underlying SST variation or on large-scale circulation variations? What is the importance of cloud-radiative feedback relative to other processes (such as surface evaporation) in regulating tropical SST?

To address these questions, we have analysed International Satellite Cloud Climatology Project (ISCCP) B3 radiance data² for the period January 1984–August 1987 (which covers almost a full El Niño cycle), and have derived changes in cirrus cloud amount (A_c), visible reflectance (R_c) and effective temperature (T_c). We used a visible-infrared threshold method to separate cirrus anvil clouds from both low or middle level clouds and small-scale convective towers³. Variations in SST and surface wind divergence were obtained from the Climate Analysis Center SST data⁴ and Florida State University surface wind data⁵, respectively.

To estimate the influence of cirrus cloud variations on broadband solar and thermal infrared fluxes at the surface, from the narrowband ISCCP radiances, we examined changes in a cloud solar index $\delta I_s = \delta[A_c(R_s - R_c)S_0]$ and in a cloud infrared index $\delta I_t = \delta[A_c\sigma(SST^4 - T_c^4)]$ caused by changes in A_c , R_c and T_c (all obtained from the ISCCP data set). $S_0 = 339.5 \text{ W m}^{-2}$ is the global annual mean incident solar flux at the top of the atmosphere, $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is Boltzmann's constant and $R_s = 6\%$ is the ocean surface reflectance.

We focus on changes in our solar index, which is defined as the difference between reflected visible radiance at the surface under clear conditions and that in the presence of cirrus (high-level) clouds. Our index isolates cases in which high-level clouds are present by excluding scenes dominated by lower-level clouds (ones which are warmer than the 500 mbar temperature). In

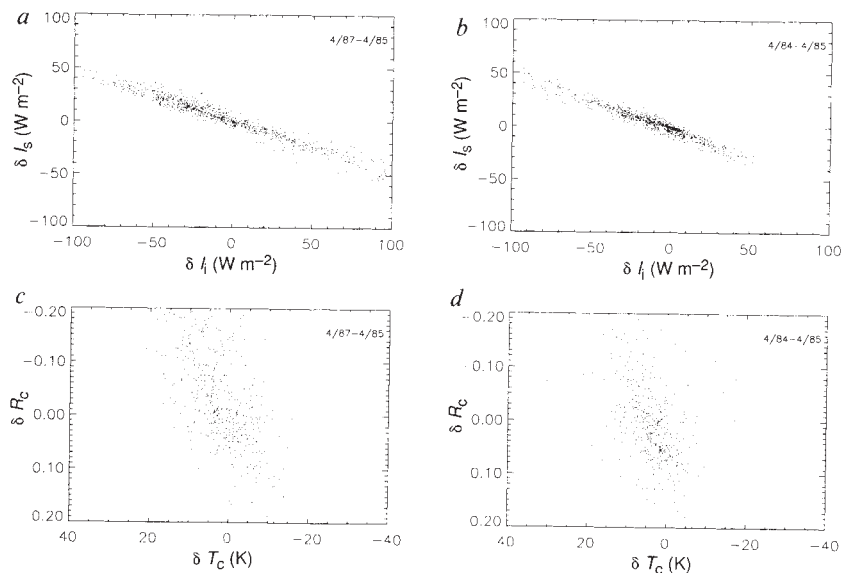
contrast, the broadband ERBE index defined by Ramanathan and Collins¹ compares clear conditions with total fluxes that include the effects of other cloud types besides cirrus. The correlation between δI_s and ERBE shortwave cloud forcing anomalies for April 1987 minus April 1985 is 0.67 (for our sample size, a 0.12 correlation is significant at the 99% level of confidence, so the probability is $\ll 1\%$ that the correlation of 0.67 occurs by chance), indicating the importance of other cloud types.

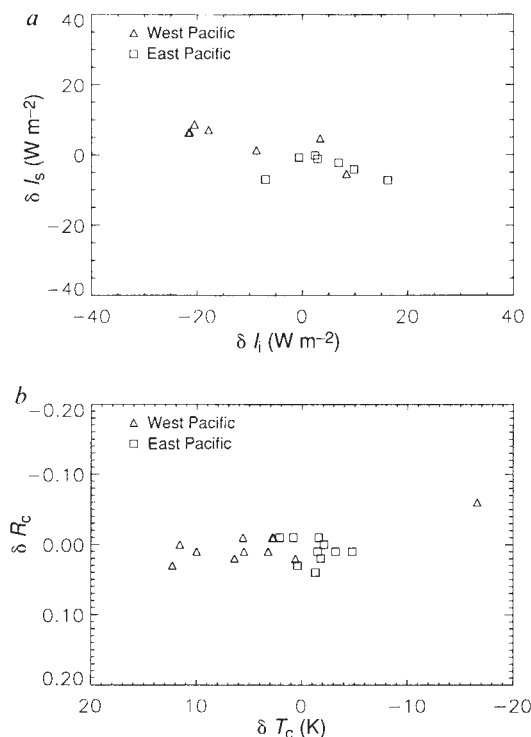
We include the infrared index for illustrative purposes only. This index is a measure of the difference between the brightness temperature (effective temperature, T_c) for scenes with high-level clouds, and the surface temperature in the absence of such clouds. As for the solar index, the high-level cirrus clouds are isolated and the lower-level ones excluded. Thus, anomalies in this index represent changes in the greenhouse effect of cirrus clouds without including the effect of the atmosphere. The Ramanathan and Collins ERBE index contrasts broadband fluxes at the top of the atmosphere in cloudy and clear conditions; interannual anomalies in their index therefore include the effects of changes not only in cirrus but also in water vapour and other cloud types. For April 1987 minus April 1985, the correlation between δI_t and ERBE longwave cloud forcing anomalies is 0.62. Note, though, that neither ERBE longwave cloud forcing nor our index represents the effect of clouds on surface fluxes, which are dominated by emission from water vapour and low-level clouds⁶.

We have correlated monthly δI_s with δI_t and δR_c with δT_c on three spatial scales: 2.5° latitude-longitude grid boxes (the scale considered in ref. 1), the tropical west (124°E – 172°W , 25°S – 25°N) and east (172°W – 90°W , 25°S – 25°N) Pacific (which averages over the Hadley circulation but separates the upwelling and downwelling branches of the Walker circulation), and the entire tropical Pacific (124°E – 90°W , 25°S – 25°N). Figure 1 shows the correlations of δI_s against δI_t and of δR_c against δT_c for 2.5° grids in the tropical Pacific for April 1987 minus April 1985 (an El Niño month minus a non-El Niño month, Fig. 1a and c), and for April 1984 minus April 1985 (two non-El Niño months, Fig. 1b and d). Figure 1a is analogous to Fig. 3b of ref. 1 except for the broader latitudinal range which does not affect the result.

The negative correlation between δI_s and δI_t means that a local increase in greenhouse effect is associated with a decrease in incoming solar radiation, which is consistent with ref. 1. Figure 1c shows that this negative correlation occurs at least in part because brightening of cirrus is associated with colder cirrus

FIG. 1 a, Scatter plot of solar against infrared flux-index differences for cirrus clouds (δI_s and δI_t , respectively, in W m^{-2}) for April 1987 minus April 1985. These are for individual $2.5^\circ \times 2.5^\circ$ latitude-longitude grid boxes in the tropical Pacific. The flux indices are defined to be positive downward. b, As in a but for April 1984 minus April 1985. c, As in a but for cirrus-cloud visible reflectance differences (δR_c , dimensionless) against cirrus-cloud infrared brightness temperature differences (δT_c , in kelvin). d, As in c but for April 1984 minus April 1985.





radiating temperatures (owing to higher cloud tops except for the thinner clouds), as assumed in ref. 1. Ramanathan and Collins interpret this correlation as meaning that the greenhouse effect "increases significantly with a warming of the ocean, and this unstable increase is effectively offset by the brighter clouds resulting from the warming."¹ But Fig. 1*b* and *d* shows that similar negative correlations occur even when the two non-El Niño months with minimal SST differences are compared. Thus, the relationship that forms the basis of the proposed thermostat mechanism is observed even when SST does not change, indicating that the cloud anomalies do not result from SST anomalies.

We repeated the analysis of Fig. 1*a* and *c*, but with data averaged over either the west or east Pacific (Fig. 2), to remove mesoscale variations, account for compensating effects within

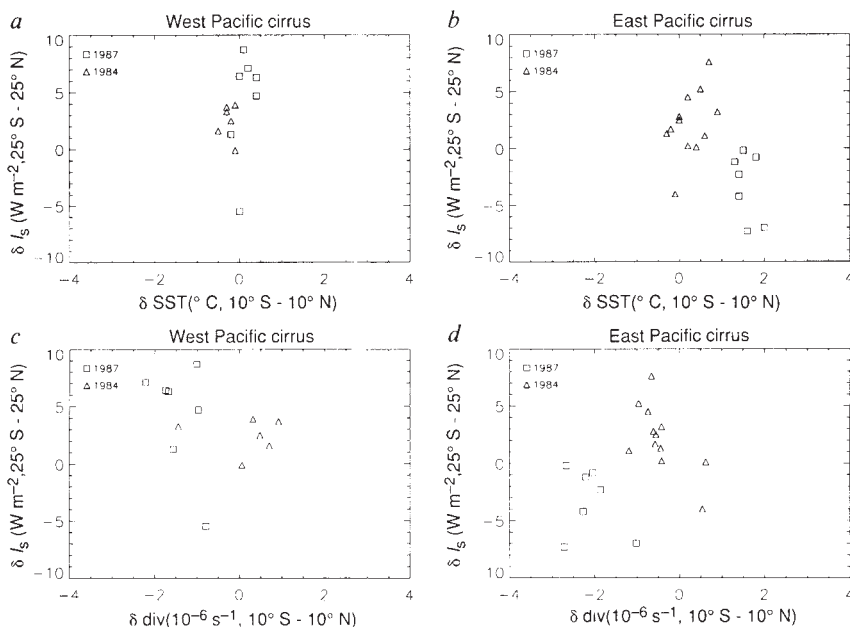
FIG. 2 *a*, Scatter plot of large-scale monthly mean I_s anomalies (δI_s) against I_i anomalies (δI_i) for 1987 minus 1985. The triangles represent anomalies averaged over the tropical west Pacific for each available month. The squares are defined similarly for data averaged over the tropical east Pacific. *b*, As in *a* but for cirrus-cloud visible reflectance anomalies (δR_c) against cirrus-cloud infrared brightness temperature anomalies (δT_c).

the Hadley cell and focus on differences between the rising and sinking branches of the Walker circulation. The negative correlation between large-scale δI_s and δI_i is less apparent (Fig. 2*a*). Incoming solar radiation slightly decreased or did not change in the east Pacific while increasing slightly or remaining constant in the west Pacific during the 1987 El Niño. Similarly, longwave-radiation trapping generally increased in the east Pacific and decreased in the west Pacific. Figure 2*b* shows that the changes in radiative indices were neither associated with any systematic anomaly in cirrus-cloud reflectance in either region nor with any cloud-top temperature anomaly in the east Pacific (where a longwave trapping anomaly is nevertheless present). Thus, El Niño perturbations of cloud radiative effects occur mainly through changes in cloud cover, not cloud top height or optical thickness. When we averaged the results over the entire tropical Pacific (not shown), we found no systematic changes caused by cirrus ($\delta I_s = -0.9$, $\delta I_i = -1.2 \text{ W m}^{-2}$), even though the basin-wide SST was warmer during the El Niño. This is because El Niño cloud changes are dominated by the geographic relocation of cirrus associated with moist convection, so that radiative variations in one area are compensated by opposite changes elsewhere.

To determine the relative impact of SST and dynamics on cirrus variability, we correlated large-scale δI_s for individual months from 1984 and 1987 with SST and surface wind divergence anomalies (Fig. 3). δI_s correlates with SST changes only in the east Pacific (Fig. 3*b*): less solar radiation is absorbed in the warmer year. In the west (Fig. 3*a*), opposite but equally large solar anomalies occur despite the absence of any SST anomaly. In both west (Fig. 3*c*) and east (Fig. 3*d*) Pacific, enhanced surface convergence occurs in the rising branch of the Hadley cell in the El Niño year. In the west, this is accompanied by an increase in incoming solar flux, whereas in the east incoming solar flux decreases.

As SST and divergence anomalies are themselves correlated in the east Pacific, it is difficult to separate their individual effects

FIG. 3 *a*, Scatter plot of large-scale monthly mean cirrus solar flux-index anomalies averaged over latitudes 25°S – 25°N (δI_s) against SST anomalies (δSST) averaged over latitudes 10°S – 10°N for two different years relative to 1985 in the west Pacific. The squares represent 1987 minus 1985 anomalies for each available month; the triangles are defined similarly for 1984 minus 1985. *b*, As in *a* but for the east Pacific. *c*, As in *a* but for solar flux-index anomalies against surface wind divergence anomalies (δdiv), with δdiv averaged over 10°S – 10°N . *d*, As in *c* but for the east Pacific.



on cirrus. But it is clear that there is no compelling evidence for an intrinsic and local relationship between SST and large-scale cirrus anvil properties in the tropics. This is especially true in the west Pacific region of maximum SSTs, where solar radiation changes as large as those in the east occur without any SST variation. In the absence of a correlation between west Pacific cloud and SST variations, SST cannot be limited to 305 K by feedbacks with cirrus clouds, as is asserted in ref. 1. We conclude that there is no evidence for the thermostat mechanism on climatically significant spatial and temporal scales.

SST is determined by a balance between ocean heat-transport and surface energy fluxes. In the tropical oceans, solar irradiance and evaporation are the dominant variable components of the surface heat flux on both monthly and interannual timescales. Ramanathan and Collins discount evaporation as a limiting factor for SST¹. To estimate the significance of latent heat flux anomalies, latent heat fluxes averaged over the entire tropical Pacific for January–August in both 1985 and 1987 were computed using a bulk parameterization model which accounts for atmospheric stability and the transition from rough air flow (moderate winds) to smooth air flow (low winds)⁷. Input parameters include SST from the Climate Analysis Center and surface wind speeds from Florida State University. The sea–air temperature difference and relative humidity are assumed to be 0.5 °C and 80%, respectively. The average evaporative flux during 1987 is 11.5 W m⁻² stronger than that for 1985. The difference

is significantly larger than δI_s (-0.9 W m^{-2}) indicating that evaporation is important for the interannual variation of surface heat flux.

Satellite observations for the 1982–1983 El Niño demonstrate that there is no significant correlation between local anomalies in surface solar irradiance and anomalous changes in SST⁸. But over a broad region south of the Equator (corresponding to large SST anomalies), there is a significant correlation between anomalous increase (decrease) in evaporation and anomalous decrease (increase) in SST. The opposite is true north of the Equator. Based on the arguments presented here, we see no evidence for the conclusion that cirrus clouds are the primary regulator of tropical SSTs. □

Received 24 January; accepted 15 June 1992.

1. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).
2. Schiffer, R. A. & Rossow, W. B. *Bull. Am. met. Soc.* **66**, 1498–1505 (1985).
3. Fu, R., Del Genio, A. D. & Rossow, W. B. *J. Climat.* **3**, 1129–1152 (1990).
4. Reynolds, W. B. *J. Clim. appl. Met.* **22**, 447–458 (1983).
5. Goldenberg, S. B. & O'Brien, J. J. *Mon. Weath. Rev.* **109**, 1190–1206 (1981).
6. Fung, I. Y., Harrison, D. E. & Lolis, A. A. *Rev. Geophys. Space Phys.* **22**, 177–193 (1984).
7. Liu, W. T., Katsaros, K. B. & Businger, J. A. *J. Atmos. Sci.* **36**, 1722–1735 (1979).
8. Liu, W. T. & Gautier, C. *J. Geophys. Res.* **95**, 13209–13217 (1990).

ACKNOWLEDGEMENTS. We thank I. Fung, M. Cane, D. Neelin and M. Ghil for constructive discussions, Y. Zhang for providing ISCCP-derived radiative fluxes and W. Tang for providing surface latent heat fluxes. This work was supported by the DOE Quantitative Links Program, the International Satellite Cloud Climatology Project, and the NASA Earth Observing System Interdisciplinary Sciences Program at JPL.

Frequency variations of the Earth's obliquity and the 100-kyr ice-age cycles

Han-Shou Liu

NASA/Goddard Space Flight Center, Geodynamics Branch, Greenbelt, Maryland 20771, USA

VARIATIONS in the Earth's orbital parameters modulate the seasonal distribution of solar radiation and thereby induce changes in the Earth's climate¹. Periodicities in the geological climate record with cycles of 100, 41 and 23 kyr have been linked with changes in eccentricity, obliquity and precession of the equinoxes, respectively^{2,3}. But although the eccentricity does vary with a 100-kyr period, the effect on the incoming solar radiation is rather weak relative to the signals from obliquity and precession variations. The 100-kyr signal in the climate record should therefore be of negligible intensity, yet it is observed instead to dominate the record. Internal, nonlinear processes within the climate system have been proposed to account for this fact^{4–14}. In contrast, I show here that variations in the frequency of the obliquity cycle can give rise to strong 100-kyr forcing of climate.

Variation of the orbital obliquity is governed by²

$$\varepsilon_{(t)} = \varepsilon + \sum_{i=1}^N A_i \cos(\omega_i t + \beta_i) \quad (1)$$

where $\varepsilon_{(t)}$ is the obliquity, and A_i , ω_i and β_i are the amplitude, mean frequency and phase angle for each component, respectively. The initial value of ε is 23.320556° and time $t=0$ refers to AD 1950. From astronomical solutions for variation in the Earth's orbital elements², the variations of the obliquity can be calculated from equation (1). The results are as follows. During the past 10⁶ years, the obliquity has oscillated for 24.57 cycles with a mean period of $T_0 = 40.7$ kyr. Two groups of longer- and shorter-period oscillations with mean period of ~42.2 kyr and 38.9 kyr (below the normal mean T_0) have occurred intermittently. The maximum values of the obliquity occurred at -9, -49, -92, -132, -171, -213, -252, -296, -334, -374, -416,

-457, -499, -538, -579, -621, -663, -702, -743, -785, -827, -869, -906, -949 and -989 kyr; and the minima at -29, -70, -112, -150, -192, -232, -274, -316, -353, -396, -436, -479, -518, -558, -600, -642, -683, -722, -764, -805, -849, -887, -927, -970 kyr. In the longer period of oscillations, the glacial phases are ~4 kyr longer than the deglacial phases. Finally, a special group of 12 obliquity oscillations has an asymmetry such that the part of the cycle that has higher tilt is shorter than the part that has lower tilt. Consequently, the period or frequency of the obliquity variation is not constant. This variation of obliquity frequency may have considerable dynamical implications because the longer glacial phases in the cycles with longer periods would produce larger ice-age ice sheets or glaciers. Most important, the special set of the 12 asymmetrical cycles, in which the deglacial phases are ~3.5 kyr shorter than the normal mean value of $T_0/2$, may accelerate the warmer summer seasons and cause rapid deglacial melting. Therefore, frequency variation in the Earth's obliquity may be an important link between the astronomical theory of climate changes and the 100-kyr glacial cycles.

The concept of the instantaneous frequency of planetary motions¹⁵ is useful for understanding the dynamical behaviour of planetary rotation^{16,17} and libration^{18,19}. Inspired by the time-frequency idea for investigation of planetary motions, I have applied this method to analyse the Earth's obliquity data and discovered that the frequency variation of the obliquity has a period of 100 kyr. To analyse the effect of these frequency variations on climate changes, we can express equation (1) as^{20,21}

$$\begin{aligned} \varepsilon_{(t)} - \varepsilon &= \sum_{i=1}^N A_i \cos(\omega_i t + \beta_i) = A_{(t)} \cos \Phi_{(t)} \\ &= A_{(t)} \cos[\omega_0 t + \phi_{(t)}] \\ &= A_{(t)} \cos \omega_{(t)} t \end{aligned} \quad (2)$$

where $A_{(t)}$ is the amplitude, $\Phi_{(t)}$ is a phase function, $\phi_{(t)}$ is phase angle, $\omega_0 = 2\pi/T_0 = 0.1543783 \text{ rad kyr}^{-1}$ and $\omega_{(t)}$ is the instantaneous frequency. Equation (2) formulates the obliquity variation for angle modulation which can be analysed in terms of frequency modulation. This is a form of angle modulation in which the instantaneous radian frequency $\omega_{(t)}$ is equal to a

constant radian frequency, ω_0 , plus a time-varying component $\Delta\omega(t) = d\phi(t)/dt$. Unlike the conventional analysis of orbital forcing which uses only Fourier transformation to produce a power spectrum, this method of frequency modulation can actually determine the dominant frequency from the instantaneous frequency spectrum and predict the time of occurrence of the main rapid melting and prolonged glaciation events.

By applying signal processing methods²², we can determine the frequency variation $\Delta\omega(t)$. The results of calculations of $\Delta\omega(t)/\omega_0$ from equation (2) are shown in Fig. 1a and b, where

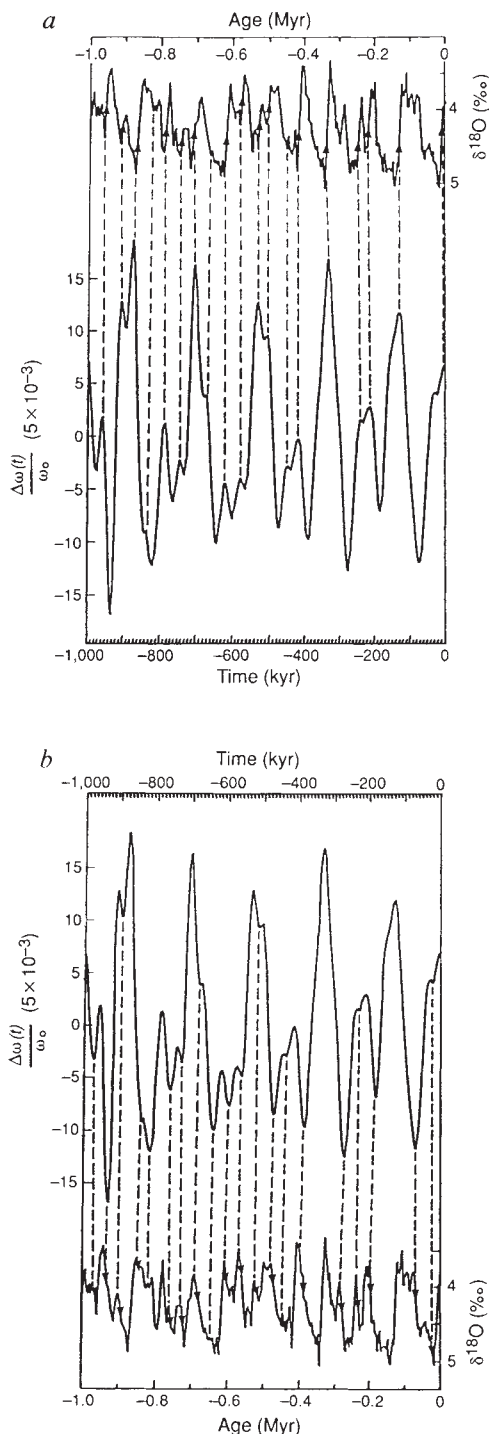


FIG. 1 a, Correlation between the maximum rates of frequency variation in the Earth's obliquity and rapid melting events in the geological climate record²³. b, Correlation between the minimum rates of frequency variation in the Earth's obliquity and large glaciation events in the geological climate record²³.

positive and negative values of $\Delta\omega(t)$ represent higher and lower frequency of obliquity variations than the mean value of ω_0 , respectively. The time intervals for the faster and slower rate of change of the obliquity are thus determined.

As shown in Fig. 1a and b, the input signal of frequency variation of the obliquity is compared with the output geological climate record²³. The time intervals at which the maximum values of frequency variation occurred are virtually related to the initiation of rapid deglacial melting, and the time intervals at which the minima occurred seem to be associated with the initiation or prolongation of principal ice ages. The spectrum of this time series is illustrated in Fig. 2. A prominent peak is seen at 100 kyr, as might be expected. In addition, there are peaks at 22, 41 and 185 kyr. The first two peaks have relatively low intensities with negligible significance. The origins of the primary 100-kyr peak and the secondary 185-kyr peak pose a problem in celestial mechanics.

The prominent 100-kyr peak in the frequency variation spectrum is of particular interest because it coincides with the anomalous 100-kyr peak in the spectrum of ice-age data. For the present problem, the origin of the 100-kyr peak in frequency variation is not the fundamental issue. The question is how any variation in the orbital elements causes nonlinear response of climate changes which may be associated with frequency variations in the obliquity. Therefore, to prove that frequency variation is responsible for rapid melting as well as large glaciation, physical processes are needed to translate these variations in obliquity frequency into a climate response.

To establish such a link, the time interval for the high obliquity and that of the high-frequency variation rate of the obliquity must be coincident. Under this condition, the Earth would enter the high-obliquity state sooner (~ 2 kyr), with an accelerated warmer summer season in northern high latitudes. This accelerated warmer season may cause rapid melting. To test this criterion, I define an index for rapid melting as an indicator for the time of the maximum obliquity coinciding with the time interval for the maximum frequency variation rate. The time indices for rapid melting, I_m , are -9, -132, -213, -334, -416, -538, -579, -621, -702, -785, -869 and -949 kyr. These indices are plotted in Fig. 3 to predict the initiation of rapid melting.

From another point of view, the rate of frequency variation of the obliquity can also be shown to be responsible for the major ice ages. The physical mechanism is as follows: to produce

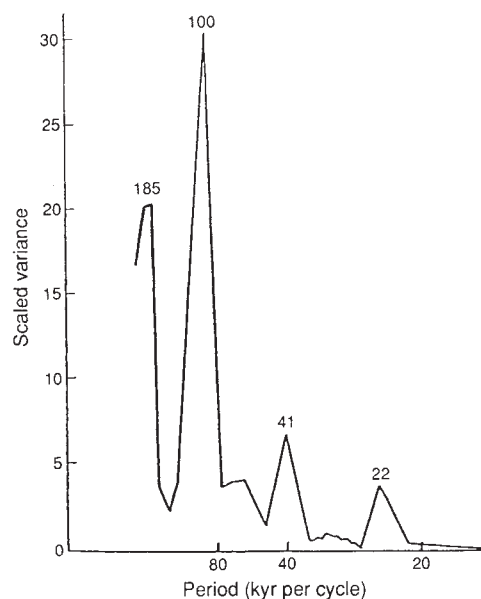


FIG. 2 Frequency variation spectrum of the obliquity.

the large ice sheets that are known to have existed, the low obliquity and low rate of frequency variation must occur in the same time interval. The Earth is constrained by this to remain the low-obliquity state with a prolonged cool summer season in northern high latitudes. The prolongation of the cool summer season (up to 3 kyr) will certainly produce large ice-age ice sheets. Again, I define an index for large glaciation as an indicator for the time of the minimum obliquity coinciding with the time interval for the minimum rate of frequency variation. The time indices for large glaciation, I_g , are -70, -192, -274, -396, -479, -558, -600, -642, -764, -849, -927 and -970 kyr. These time indices are also plotted in Fig. 3 to predict the initiation or prolongation of large glaciation.

The agreement between the indices for rapid melting and large glaciation (I_m , I_g) found here, and the events of rapid melting and large glaciation in the geological record of climate changes as shown in Fig. 3, seems to verify the theory that the variation of the obliquity frequency is a principal forcing function for the 100-kyr ice-age cycles. Of course, the conditions for I_m and I_g are constrained by other orbital elements such as eccentricity and precession cycles. For instance, the incorrect and missing prediction of the initiation of large ice ages at -320 kyr and -680 kyr is attributable to the result of the amplitude modulation of the precession by the eccentricity²⁴.

The importance of the 100-kyr frequency variation cycles in the excitation of the 100-kyr ice-age cycles has been examined by means of cross-correlation in the time domain. A correlation study shows that the main events of the rapid melting and prolonged glaciation in the 100-kyr ice age cycles lag behind the indices I_m and I_g in the 100-kyr obliquity-related cycles by ~2.4 kyr and 3.8 kyr, respectively.

I have done a further series of analyses of orbital data to test the coupling effect of the time indices (I_m , I_g) and the eccentricity. For the main melting events, the time indices I_m coincide with the maximum values of variation of the eccentricity. For the main glaciation events, however, the time indices I_g are not related to the minimum values of the eccentricity. These findings may provide a way of discriminating between the obliquity-related 100-kyr period and the 100-kyr cycle in the eccentricity. According to principles of celestial mechanics²⁵, variations in the Earth's obliquity are produced by a coupling between the motion of its orbital plane due to the gravitational perturbations of the other planets and the precession of the spin axis which results from the solar torque exerted on the Earth's equatorial bulge. Variations in the magnitude of this solar torque are determined by the instantaneous distance from the Earth to the Sun, which is governed by the eccentricity. Therefore, the frequency variation of the obliquity may be partially controlled

by the eccentricity in an intriguing way. It can be argued²⁶ that the maximum rates of frequency variation of the obliquity can occur only when the eccentricity is high. The minimum rates of frequency variation, however, can occur regardless of the magnitude of the eccentricity. If this interpretation is correct, the 100-kyr cycle in eccentricity could not be the forcing mechanism for the formation of the 100-kyr glacial cycles; it can be considered only as one of the parameters that help to produce the 100-kyr cycle in rapid melting.

Finally, the dilemma of the 100-kyr ice-age cycle during the past million years is further complicated by the observation that before ~1 million years ago the 100-kyr cycle in global climate was far less important. The effect of frequency variations in the obliquity on climate changes before 1 million years ago is being tested. □

Received 24 October 1991; accepted 4 June 1992.

1. Milankovitch, M. *Konigl. Serb. Akad. Beograd. Spec. Publ.* **132**, (1941). (Engl. transl. Israel Program for Scientific Translations, Jerusalem, 1969).
2. Berger, A. L. *Nature* **269**, 44–45 (1977).
3. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
4. Imbrie, J. & Imbrie, J. Z. *Science* **207**, 943–953 (1980).
5. Pollard, D. *Nature* **272**, 233–235 (1978).
6. Pollard, D. *Nature* **296**, 334–338 (1982).
7. Oerlemans, J. *Nature* **287**, 430–432 (1980).
8. Peltier, W. R. *Adv. Geophys.* **24**, 2–146 (1982).
9. Shackleton, N. J. *et al. Nature* **307**, 620–623 (1984).
10. Ruddimen, W. F. & McIntyre, A. *Geol. Soc. Am. Bull.* **95**, 381–396 (1984).
11. Saltzman, B., Hansen, A. R. & Maasch, K. A. *J. Atmos. Sci.* **41**, 3380–3380 (1984).
12. Weertman, J. *Nature* **261**, 17–29 (1976).
13. Hyde, W. T. & Peltier, W. R. *J. Atmos. Sci.* **44**, 1351–1374 (1987).
14. Birchfield, G. E., Weertman, J. & Lunde, A. T. *J. Atmos. Sci.* **39**, 71–87 (1982).
15. Liu, H. S. & O'Keefe, J. A. *Science* **150**, 1717 (1965).
16. Laslett, L. J. & Sessler, A. M. *Science* **151**, 1384–1385 (1966).
17. Jefferys, W. H. *Science* **152**, 201–202 (1966).
18. Liu, H. S. & Chao, B. F. *Geophys. J. Int.* **106**, 699–702 (1991).
19. Chao, B. F., Dong, D. N., Liu, H. S. & Herring, T. A. *Geophys. Res. Lett.* **18**, 2007–2010 (1991).
20. Black, H. S. *Modulation Theory*, 363 (Van Nostrand, London, 1975).
21. Ingard, K. U. *Fundamentals of Waves and Oscillations*, 591 (Cambridge Univ. Press, New York, 1990).
22. Brook, D. & Wynne, R. J. *Signal Processing*, 308 (Edward Arnold, London, 1988).
23. Shackleton, N. J., Berger, A. L. & Peltier, W. R. *Trans. R. Soc. Edinb. Earth Sci. Rev.* **81**, 251 (1990).
24. Vernekar, A. D. *Meteorological Monographs* **12** (Am. met. Soc., Boston, 1972).
25. Ward, W. R. *Science* **181**, 260–262 (1973).
26. Bills, B. G. *J. geophys. Res.* **95**, 14137–14153 (1990).

ACKNOWLEDGEMENTS. I thank B. G. Bills and B. F. Chao for discussions and C. Wade for computation.

Evidence for basal marine ice in the Filchner–Ronne ice shelf

H. Oerter*, J. Kipfstuhl*, J. Determann*, H. Miller*,
D. Wagenbach†, A. Minikin† & W. Graf‡

* Alfred-Wegener-Institut für Polar- und Meeresforschung,
Postfach 120161, D-W2850 Bremerhaven, Germany
† Institut für Umweltphysik, Universität Heidelberg,
Im Neuenheimer Feld 366, D-W6900 Heidelberg, Germany
‡ GSF-Institut für Hydrologie, Ingolstaedter Landstrasse 1,
D-W8042 Neuherberg, Germany

THE Filchner–Ronne ice shelf, which drains most of the marine-based portions of the West Antarctic ice sheet, is the largest ice shelf on Earth by volume. The origin and properties of the ice that constitutes this shelf are poorly understood, because a strong reflecting interface within the ice and the diffuse nature of the ice–ocean interface make seismic and radio echo sounding data difficult to interpret^{1,2}. Ice in the upper part of the shelf is of meteoric origin, but it has been proposed^{2–5} that a basal layer of saline ice accumulates from below. Here we present the results of an analysis of the physical and chemical characteristics of an ice core drilled almost to the bottom of the Ronne ice shelf. We observe a change in ice properties at about 150 m depth, which we ascribe to a change from meteoric ice to basal marine ice. The basal ice is very different from sea ice formed at the ocean surface,

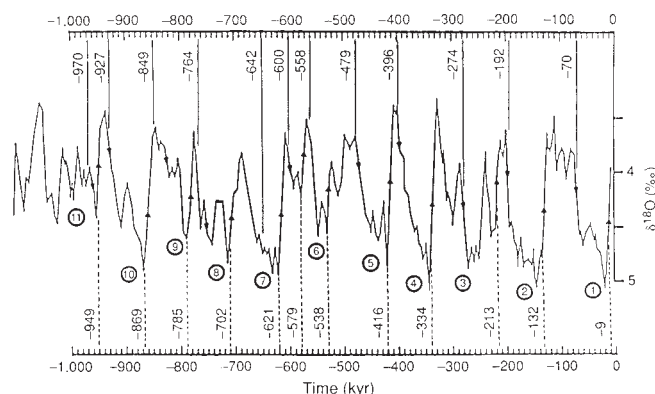


FIG. 3 Time indices for rapid melting (I_m) and large glaciation (I_g), calculated from the input obliquity signal, predict major events of rapid melting and large glaciation in the output geological record of climate changes²³. (Solid lines represent I_g ; dotted lines represent I_m ; numbers in circles represent the sequence of glacial stages.)

and we propose a formation mechanism in which ice platelets in the water column accrete to the bottom of the ice shelf.

During austral summer 1989–90 (RV *Polarstern* cruise ANT VIII/5⁶), roughly 30 km inland from the front of the Ronne ice shelf, Antarctica (Fig. 1), where the total ice shelf thickness is 239 m (N. Blindow, personal communication), the ice core B13 was recovered from the surface to a depth of 215 m. We report here electrolytic conductivity measurements and oxygen isotope ratios of drill chipping samples in the liquid state. More detailed isotopic and chemical analyses as well as ice fabric studies are still under way.

The top 152.8 m of the core consisted of firm and bubbly ice of meteoric origin. Below this depth the ice changed at a sharp horizon to clear, bubble-free ice which contained layers of particles, and was clearly recognizable by visual inspection. The boundary corresponds to the radio-echo horizon, erroneously interpreted as the base in some earlier surveys¹, but later found to be an internal reflector². The textural, isotopic and chemical data presented here provide evidence that this basal ice layer originates from sea water. But the process of its formation is substantially different from that involved in forming sea ice at the sea surface. To distinguish the two, we introduce the term 'marine shelf ice'.

Thin sections of two core segments (Fig. 2) display the general textural patterns to be found in the marine shelf ice. Although the smallest grains and the highest diversity appear in the uppermost and oldest parts (for example, at 153.7 m, Fig. 2a), grain sizes are several times larger in the deeper and younger layers at 190 m (Fig. 2b). Horizontal layering, which usually occurs in conjunction with particle inclusions and obviously prohibits crystal growth, is only found in the upper parts of the marine shelf ice. Preliminary microscopic analyses of the inclusions revealed mainly clastic particles, and some single diatom frustules as well as remnants of radiolarians. Although the marine shelf ice originates from sea water, its properties are distinctly different from those of sea ice. Specific sea-ice characteristics such as brine drainage channels, bubbles, pronounced c-axis alignments or an evenly spaced substructure⁷ are completely absent from our samples.

The $\delta^{18}\text{O}$ values of core B13 (Fig. 3a) steadily decrease from about -26% near the surface to -33.5% at a depth of ~ 135 m. This is in good agreement with the isotope content found in surface samples at locations further south along the flow line, from where the ice found in deeper parts of the core originates.

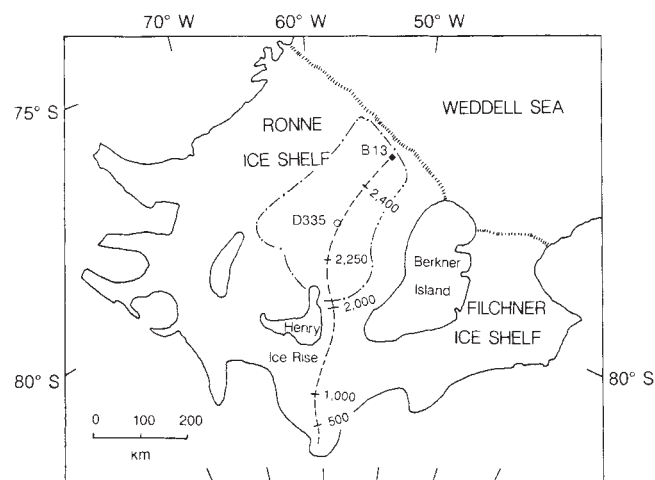


FIG. 1 Map of Filchner-Ronne ice shelf showing the location of the ice-core drill hole B13 in 1990, the hot-water drill hole at point D335 from 1986³, and an ice flow line through the location of the ice-core B13 (---). Travel times of the ice shelf to the drill location are indicated. The assumed location² of the bottom layer of marine shelf ice is also shown (— · — · —).

But the $\delta^{18}\text{O}$ values decrease more rapidly between depths of 135 m and 152 m, reaching a minimum value of -40% as the transition horizon is approached. The rapid decrease suggests that this portion of the ice core was originally deposited as inland ice of higher elevation. As ice thickness at the grounding line (inland margin) exceeds 1,400 m, a reduction to ~ 17 m close to the ice front provides evidence for strong basal melting between the grounding line and the area where basal accumulation starts, at $\sim 80^\circ\text{S}$ ^{2,5}. Without basal melting one would expect that the ice reaching the ice front would be >100 m thick (assuming vertical thinning rates derived from an ice shelf flow model⁵ along a flow line through the drilling site). Thus the $\delta^{18}\text{O}$ values of ice core B13 confirm basal melting beneath the deep parts of the Ronne ice shelf; this has been postulated previously from mass balance studies⁹ and from modelling the ocean circulation under the ice shelf^{9–11}. As explained later, we believe that basal melting at the grounding line is essential for the formation of marine shelf ice.

The depth profiles of $\delta^{18}\text{O}$ values and the electrolytic conductivity show very pronounced shifts at the visually recognized boundary of the two ice bodies (Fig. 3a and b). Within ~ 2 cm, the $\delta^{18}\text{O}$ values (given as the deviation from the Vienna standard mean ocean water, V-SMOW) change from -40 to $+2\%$ and define the boundary between meteoric ice and marine shelf ice. The $\delta^{18}\text{O}$ value of $+2\%$ indicates that the marine shelf ice is frozen from sea water that has been enriched with glacial meltwater. The electrolytic conductivity of the marine shelf ice (Fig. 3b) corresponds to salinities of $<0.1\%$. This is one order of

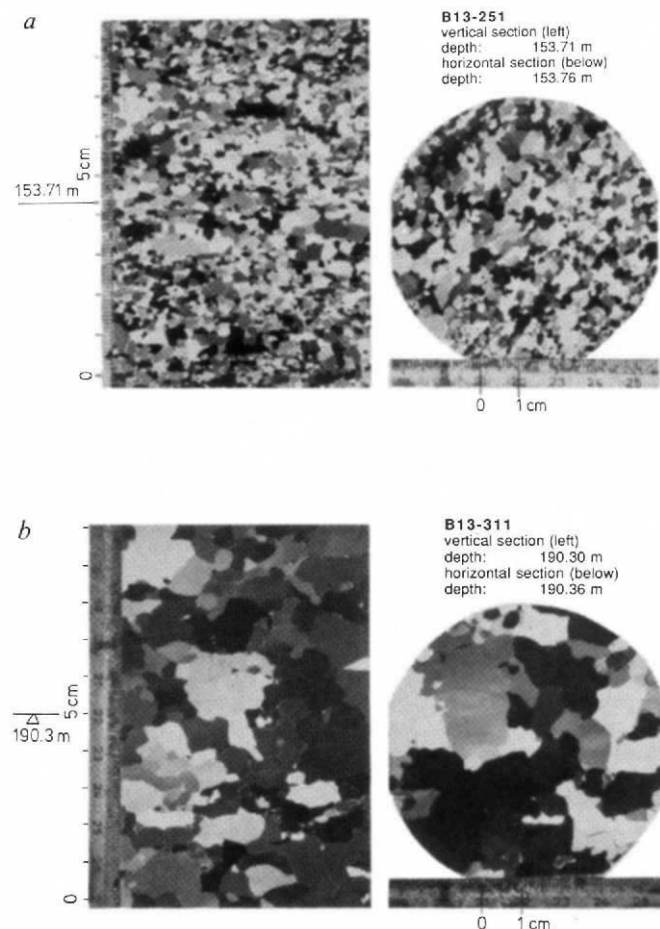


FIG. 2 Thin sections of marine shelf ice, recovered with ice-core B13 in 1990 at the central part of Ronne ice shelf from two different depths: a, 153.7 m, close to the topmost layer of the marine shelf ice; b, 190.3 m, the bottom part of the ice core.

magnitude above the values recorded for meteoric ice, but it is much lower than the typical salinity of sea ice ($\sim 3\text{--}10\%$, ref. 7). The several-hundred-years-old congelation ice, frozen to the base of Ross ice shelf and recovered with the J9 ice core, still had salinities of 2–4% (ref. 12).

Detailed chemical analysis has been done along the entire core of the marine shelf ice on chippings, and in 1-cm depth resolution on two selected core sections. The chloride concentration decreases from 40 p.p.m. in the uppermost marine shelf ice layers to 10 p.p.m. below 180 m depth, which is consistent with the electrolytic conductivity record. The effective salt distribution coefficient (the ratio of the salt concentration in the ice to that of the bulk sea water) is therefore $\sim 10^{-3}$, suggesting that the build up of the marine shelf ice is driven by ice-forming mechanisms that are highly effective in rejecting common seawater ions. Accordingly, the ionic ice composition shows significant deviations from the corresponding seawater proportion: whereas SO_4^{2-} and Mg^{2+} are depleted with respect to Cl^- , Na^+ is enriched. This means that temperature-dependent precipitation of cryohydrates (such as mirabilite, $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$; ref. 13) which is responsible for the chemical fractionation of sea ice and sea-ice brine, is an unlikely cause of this phenomenon. Moreover, the fractionation occurring in the marine shelf ice increases with depth. The sulphate/chloride mass ratio deviates from the seawater ratio of 0.14 by -5% in 155 m depth, and by -42% in 214 m depth. The corresponding deviations of the sodium/chloride ratio are $+3.5\%$ and $+14.5\%$, and for the magnesium/chloride ratio -10% and -43% , respectively. The decrease with depth of the electrolytic conductivity (the salinity) and the stronger chemical fractionation in the deeper layers both suggest a decrease in the formation rate for deeper (younger) parts of the marine shelf ice body.

The general pattern of both strong basal melting near the inland margin and basal accumulation in the central Ronne ice shelf can be explained by a large-scale ice pump¹⁴ working

between the grounding line and the central part of the ice shelf (Fig. 1). This deep thermohaline convection cell in the sub-ice-shelf cavity is driven by the pressure dependence of the freezing point. At the grounding line the ice-shelf draft exceeds 1,000 m, leading to an *in situ* freezing point between -2.5 and -3°C . Basal melting is caused when water masses which are above this temperature come into contact with the ice-shelf base. Being diluted by meltwater, they then become colder but buoyant and ascend along the ice-shelf base towards the north. On approaching the shallower areas north of Henry ice rise, the water becomes supercooled and tends to release ice crystals that may be similar to the ice platelets trawled offshore of the Filchner ice shelf¹⁵. The ice formation is associated with the release of latent heat and salt, leading to a warmer but denser water mass which sinks and drives a deeper counter current, transporting heat to the grounding line. Recently this hypothesis has been corroborated by profiles of temperature and salinity in the western Ronne ice shelf¹⁶.

The formation of marine shelf ice might then be explained by the following process. On their ascent through the water column the platelets scavenge suspended particles (which are found in the ice core) and initially form a slushy layer some tens of metres thick^{3,16}. Further compaction by buoyant forces and metamorphic processes, enhanced by temperatures close to the freezing point, might then create the marine shelf ice. The observed very low salt concentration and the chemical fractionation could in principle be the result of both processes: the formation of freely floating ice crystals in the water column and their recrystallization during the evolution of the marine shelf ice. As the heat conduction of the ice shelf is very low, the formation of congelation ice is limited to a few centimetres per year¹². Therefore it cannot explain the formation of the marine shelf ice.

An almost identical type of marine shelf ice found at the base of an overturned iceberg in the eastern Weddell sea¹⁷ indicates that such ice does not only exist in the Filchner–Ronne ice shelf. Moreover, the $\delta^{18}\text{O}$ profile of the G1 ice core from the Amery ice shelf¹⁸ displays a similar pattern to the $\delta^{18}\text{O}$ profile from ice core B13. In both cases a three-layered structure is found: meteoric ice deposited on the ice shelf, inland ice from the continent and marine shelf ice originating from sea water. Note that in both $\delta^{18}\text{O}$ profiles, very low $\delta^{18}\text{O}$ values typical for ice-age ice are absent, showing that the deepest layers of the inland ice do not survive at the ice-shelf bottom.

As inferred from an ice-shelf model⁵, 40 km^3 of marine shelf ice have to be formed each year to keep the central part of the Ronne ice shelf in a steady state. Comparing this, for example, with the $90\text{ km}^3\text{ yr}^{-1}$ of surface accumulation for the entire Filchner–Ronne ice shelf (assuming a mean precipitation rate of 0.19 m of ice per year⁸), demonstrates the important role of the marine shelf ice within the mass balance and dynamics of the Filchner–Ronne ice shelf. $\square$

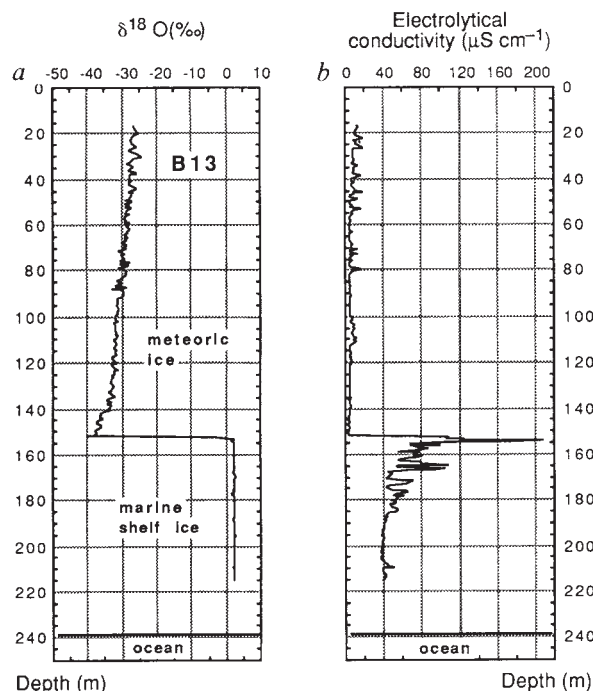


FIG. 3 Ice-core B13, Ronne ice shelf, 1990. *a*, $\delta^{18}\text{O}$ profile measured at drill chipping samples. *b*, Profile of the electrolytic conductivity, measured on melted drill chipping samples. The boundary between meteoric ice and marine shelf ice is displayed with the sharp increase of $\delta^{18}\text{O}$, and electrolytic conductivity. The bottom of the ice shelf was at $\sim 239\text{ m}$ (N. Blindow, personal communication).

Received 24 January; accepted 5 April 1992.

1. Robin, G. de Q. *et al.* *Nature* **302**, 582–586 (1983).
2. Thyssen, F. *Ann. Glaciol.* **11**, 173–179 (1988).
3. Engelhardt, H. & Determann, J. *Nature* **327**, 318–319 (1987).
4. Lange, M. A. & MacAyeal, D. R. *J. geophys. Res.* **91**, 10457–10462 (1986).
5. Determann, J. *Antarct. Sci.* **3**, 187–195 (1991).
6. Miller, H. & Oerter, H. (eds) *Berichte zur Polarforschung* **86** (1991).
7. Weeks, W. F. & Ackley, S. F. *U.S. Army Cold Regions Res. Eng. Lab. Monogr.* 82-1 (Hannover, 1982).
8. Graf, W., Moser, H., Oerter, H., Reinwarth, O. & Stichler, W. *Ann. Glaciol.* **11**, 23–31 (1988).
9. Jenkins, A. J. & Doake, C. S. M. *J. geophys. Res.* **96**, C1, 791–813 (1991).
10. Hellmer, H. H. & Olbers, D. J. *Antarct. Sci.* **1**, 325–336 (1989).
11. Hellmer, H. H. & Olbers, D. J. *Antarct. Sci.* **3**, 433–442 (1991).
12. Zotikov, I. A., Zagorodnov, V. S. & Raikovskiy, J. V. *Science* **207**, 1463–1465 (1980).
13. Assur, A. *Nat. Acad. Sci./Nat. Res. Council, Arctic Sea Ice, Publ.* 598, 106–138 (Washington DC, 1958).
14. Lewis, E. L. & Perkin, R. G. *J. geophys. Res.* **91**, 11756–11762 (1986).
15. Dieckmann, G., Rohardt, G., Hellmer, H. H. & Kipfstuhl, J. *Deep Sea Res.* **33**, 141–148 (1986).
16. Nicholls, K. W., Makinson, K. & Robinson, A. V. *Nature* **354**, 221–223 (1991).
17. Kipfstuhl, J., Dieckmann, G., Oerter, H., Hellmer, H. H. & Graf, W. *J. geophys. Res.* (in the press).
18. Morgan, V. I. *Nature* **238**, 393–394 (1972).

Post-depositional mobility of platinum, iridium and rhenium in marine sediments

Debra C. Colodner^{*‡}, Edward A. Boyle^{*},
John M. Edmond^{*} & John Thomson[†]

^{*} Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, E34-201, Cambridge,
Massachusetts 02139, USA

[†] Institute of Oceanographic Sciences Deacon Laboratory, Brook Road,
Wormley, Godalming, Surrey GU8 5UB, UK

THE discovery of high concentrations of iridium in Cretaceous/Tertiary boundary sediments engendered the hypothesis¹ that a meteorite collided with the Earth 65 million years ago, coincident with the mass extinction that occurred at that time. Iridium spikes of various magnitudes have subsequently been reported at more than 10 other extinction horizons^{2–11}. It has been suggested, on the other hand, that geochemical processes might create or modify many of these spikes^{5,11–14}, but a critical evaluation of these suggestions has been hindered by incomplete understanding of low-temperature iridium geochemistry. Other platinum-group elements (Ru, Rh, Pd, Re, Os, Pt, Au) are often found to be associated with Ir spikes, and inter-element ratios have been used to assess the cosmic or terrestrial nature of the enrichments^{5,7,9,28,29}; but the geochemical influences on these relative abundances are also poorly constrained. Here we describe analyses of recent abyssal marine sediments which allow us to characterize the behaviour of Pt, Re and Ir during early diagenesis. These elements are redistributed by changes in sedimentary redox conditions. Such changes can probably account for many of the small platinum-group-element spikes found in the geological record, and may render ambiguous attempts to interpret inter-element ratios.

Platinum-group-element (PGE) anomalies in sediments may result from deposition of an enriched source of the elements (such as cosmic or mantle-type material) and/or from changes in the sedimentary environment. The latter possibility has been suggested in several previous studies, which have proposed the following mechanisms for PGE enrichment: precipitation with bacterial iron oxides^{11,12} or sulphides¹³, concentration from sea water into reducing sediments^{5,9} and removal from ground water at redox boundaries within the sediment column¹⁴. Here we examine Pt, Ir and Re migration in response to redox changes driven by degradation of organic matter in marine sediments. Recent studies^{15–17} of the marine geochemistry of Pt and Ir show that both elements are relatively enriched in manganese nodules, with concentrations of up to 900 parts per 10⁹ (p.p.b.) Pt and 7 p.p.b. Ir (compared with ~3 p.p.b. Pt and 0.2 p.p.b. Ir in pelagic clays). It seems that Pt and Ir are incorporated into hydrogenous (seawater-derived) Fe–Mn oxyhydroxide minerals, and that these phases might dominate the distribution of these elements in sediments. Conversely, Re is known¹⁸ to be accumulated from sea water under reducing conditions, reaching concentrations of ~50 p.p.b. in organic-rich sediments, compared with ~0.1 p.p.b. in oxic sediments.

We determined bulk-sediment Pt, Ir and Re distributions in abyssal-plain sediments from the eastern foothills of the Mid-Atlantic Ridge (core T1) and the Nares (P1) and Madeira abyssal plains (T2). In these regions, pelagic sediments (accumulating at a rate of 0.2–0.5 cm kyr⁻¹) are interrupted sporadically by turbidite units which originate on the continental slopes. The turbidites are extremely distal, as revealed by near-constant mineralogy and major-element chemistry, and are relatively rich

TABLE 1 Pt, Ir and Re data for Atlantic abyssal sediments

Core description	Depth (cm)	Pt (p.p.b.)	Ir (p.p.b.)	Re (p.p.b.)
T1	105–106	1.0	0.03	0.3
turbidite	110–111	1.2	0.02	0.7
core 11805 #1 K	111–112	1.2	0.02	0.4
25° 39' N	112–113	0.9	0.05	0.3
30° 57' W	113–114	1.5	0.03	1.0
6,050 m	114–115	2.0, 2.1	0.04, 0.02	6.6, 6.1
	115–116	2.5	0.04	4.8
	116–117	2.8	0.03	6.6
	117–118	2.5	0.03, 0.04	11.5
	118–119	2.4, 2.5	0.04, 0.05	10.5
	119–120	2.6	0.04	11.0
	120–121	1.6, 1.8	0.06, 0.04	15.4
	121–122	1.6	0.05	10.6
	130–131	1.0, 1.0	0.04	14.1
	140–141	0.8, 1.1	0.04, 0.06	14.0
	150–151	1.1	0.04	
P1	1580	3.7	0.11	0.05
pelagic halo	1582	2.7	0.11	0.05
core 52–2	1583	3.4	0.17	0.05
23° 22.3' N	1594	1.7	0.07	0.04
63° 00.7' W	1595	1.9	0.06	0.03
5,878 m	1596	1.2	0.06	0.04
	1597	2.5	0.03	0.07
	1599	3.1	0.07	0.07
	1600	2.0	0.08	0.07
	1602	1.9	0.10	0.06
	1603	4.2	0.09	0.07
	1604	4.6	0.16	0.06
	1605	5.1	0.15	0.06
	1606	5.2	0.17	0.07
	1607	5.2	0.17	0.08
	1609	3.9	0.16	0.07
For comparison: C1 chondrites ²⁷		953	473	37

Pt, Ir and Re were determined by isotope dilution–inductively coupled plasma mass spectrometry as described in ref. 25: 0.5–1 g dry sediment was dissolved in acid (HF–HNO₃–HCl) in Teflon bombs in a microwave oven, and Pt, Ir and Re were preconcentrated by anion exchange. Samples were introduced to the ICP–MS (VG Plasmaquad) using a flow-injection valve with a 250 µl loop. Results are reported on a whole-sediment basis (not carbonate-free).

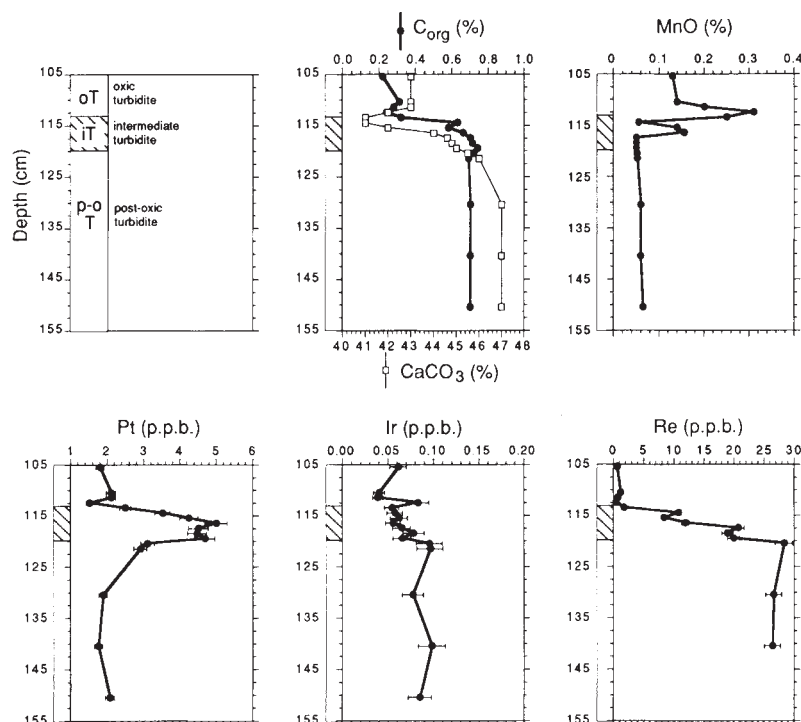
in organic carbon (up to 1%). These sediments are ideally suited to a study of post-depositional mobility because of the instantaneous nature of turbidite emplacement and the extensive array of chemical data that already exists^{19–22}. Following turbidite deposition, diagenetic alteration of both the turbidite and the underlying pelagite takes place. Diffusive penetration of bottom-water oxygen and nitrate into the turbidite leads to the oxidation of organic matter above a progressively deepening oxidation front. This results in the formation of a partially oxidized turbidite, with oxic sediments overlying sub-oxic (containing neither oxygen nor sulphide) sediments. The arrival of the turbidite also isolates the former pelagic surface from bottom-water oxidants. The oxidation of any labile organic matter present within the pelagite proceeds with the reduction of Fe–Mn oxyhydroxides, resulting in a 'bleached' interval (a reduction halo²⁰) at the top of the pelagic unit.

Many transition metals are redistributed in response to these redox changes in the sediments^{19–24}. For elements with several possible valence states, oxidized and reduced forms may show different susceptibilities to adsorption and/or precipitation reactions. Additionally, metals may be adsorbed by or precipitated with phases that are specific to oxic or anoxic sediments, such as Fe–Mn oxyhydroxides or sulphides. A combination of these factors will determine the distribution of the element between pore waters and solid phases, and hence elemental mobility.

We measured Pt, Ir and Re distributions in four cores; two contain partially oxidized turbidite units and the other two

[‡] Present address: Geochemistry Department, Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA.

FIG. 1 Measurements of organic carbon (C_{org}), $CaCO_3$, MnO, Pt, Ir and Re for core 11805 (T1), containing a partially oxidized, calcareous turbidite. The uppermost 50 cm of the core consists of pelagic brown clay which overlies a turbidite stratum at least 11 m thick. The profiles shown here are entirely within the turbidite unit. Pore-water data indicate that oxygen is present in the pore waters to ~120 cm, and that sulphate reduction is negligible in this section. The turbidite is believed to have been emplaced 330 kyr ago. The oxic and sub-oxic sections of this turbidite are separated by a transition zone of intermediate colour from 113 to 120 cm. The upper, oxidized portion has suffered some calcite dissolution; Mn, Pt, Ir and Re results are therefore plotted on a carbonate-free basis. Precision of the data ($\pm 6\%$ for Pt, $\pm 20\%$ for Ir and $\pm 5\%$ for Re) was determined by processing some samples in duplicate and repeated runs of other sediments²⁵. MnO and C_{org} data are from Thomson *et al.*²¹.



contain partially reduced pelagic sections. The conclusions that we draw from the two turbidite sections are essentially identical, as are those from the two pelagic sections²⁵, and we therefore present results for only one example of each (T1 and P1) in Table 1. Each core shows variations in transition metal concentrations which are not related to primary mineralogy and are best explained by post-depositional mobility¹⁹⁻²⁴. Results for Mn, Co, Cu, V and I are presented in Figs 1-3 as an aid to interpretation of the Pt, Ir and Re results.

The turbidite in core T1 is at least 11 m thick and was emplaced approximately 330 kyr ago. The oxidized and unoxidized sections of this turbidite are separated by a transition zone of intermediate colour, through which the concentration of organic carbon increases. Pt, Ir and Re show markedly different responses to the partial oxidation of this turbidite (Fig. 1). Pt is enriched in a broad peak toward the base of the transition zone. Ir variations are more poorly established, but there is some indication of Ir loss from the oxic section and transition zone, with possible enhancement (one point) at the Mn peak. Re is enriched in the unoxidized, organic-rich turbidite, but has been lost from the upper, oxidized section, with a profile similar to that of organic carbon.

An expanded version of the transition region, along with I, Cu and V distributions, is presented in Fig. 2a. The Pt peak

occurs within a sequence of elemental peaks surrounding the redox boundary, this sequence (from top to bottom) being Mn, I, Pt, Cu and V. This is also seen in a second core (T2, Fig. 2b) with the same ordering of elements in the redox gradient (there is no Mn peak in core T2, merely enrichment in the oxic section). Whereas Mn and I diffuse out of the unoxidized portion of the turbidite and become immobilized after crossing the redox boundary into the upper, oxidized section, Cu and V are remobilized continuously as the turbidite is oxidized and become fixed below the redox boundary²¹. The sampling intensity for Pt above and below the peaks is insufficient to establish whether the oxic or sub-oxic portions of the turbidites have higher Pt concentrations. For this reason, and because of the lack of pore-water data, it is not possible to determine definitively whether Pt has migrated to the boundary from above or from below.

The effects of reduction of oxic hydrogenous phases may be examined in core P1 (Fig. 3). In this sediment, a reduction halo (~9 cm thick) has developed below a thin turbidite. Mn and Co have been removed from this interval, with Cu exhibiting intermediate behaviour, increasing gradually toward the bottom of the halo. Some of the Co released from the halo has been fixed in the pelagic sediment²⁰. Pt and Ir in this core have also been partly removed from the halo, with profiles showing similarities to both Co and Cu. Roughly 50% of the

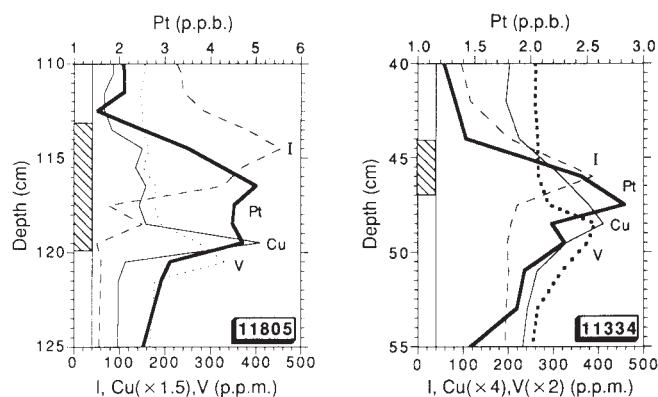
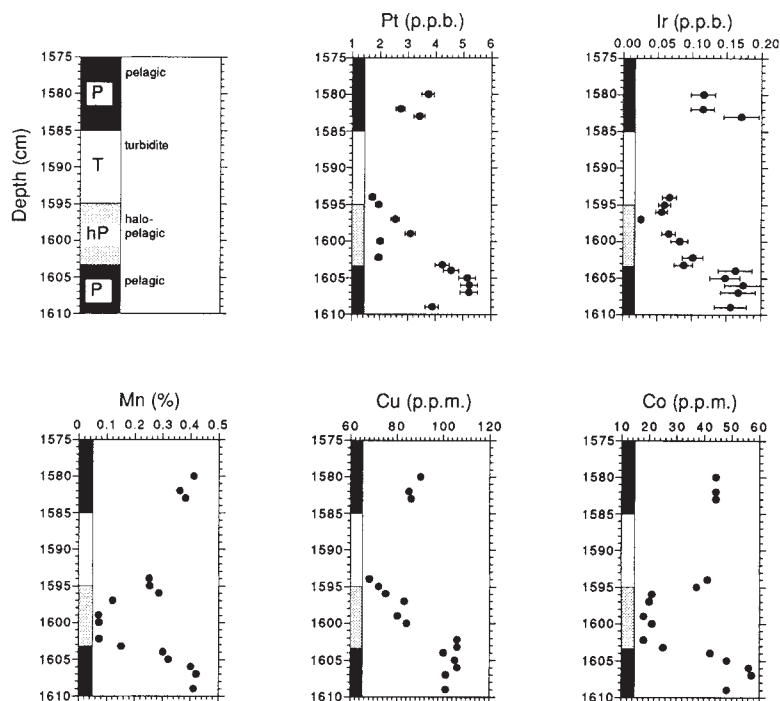


FIG. 2 Distributions of I, Pt, Cu and V surrounding the oxic-sub-oxic transition zone in (left) core 11805 (T1) and (right) core 11334 (T2). The relative positions of the elements in the redox gradient are the same in the two cores. Cu concentrations in both cores and V concentrations in T2 have been multiplied by an arbitrary factor so that the four elements could be plotted on the same graph. I, Cu and V concentrations are from Thomson *et al.*²¹ and J. Thomson (unpublished data).

FIG. 3 Measurements of Pt, Ir, Mn, Cu and Co for core 52-2 (P1), consisting of pelagic sediments interrupted by a turbidite from 1,585 to 1,595 cm. A reduction halo has developed below the turbidite (1,595–1,603.5 cm). The age of this section is believed to be 1.5–2 myr. Organic carbon comprises <0.15% of the pelagic clays, based on data from a nearby core²⁰ (ESOPE core 48; 22° 54.9' N, 63° 26.7' W). As the main goal of this part of the study was to compare the reduction halo to the unaltered pelagic sediment, samples from the turbidite (1,585–1,595 cm) have not yet been analysed. Precision of the data ($\pm 6\%$ for Pt, $\pm 15\%$ for Ir and $\pm 15\%$ for Re) was determined by repeated runs on other sediments²⁵. Mn, Cu and Co data are from Thomson *et al.*²⁰



Pt and 40% of the Ir originally present has been mobilized. Re concentrations are low throughout this oxic pelagic core and are the same within the halo and unaltered sediment (Table 1).

Pt, Ir and Re mobility in these sediments is contingent on their significant hydrogenous enrichments. Development of these enrichments depends both on the ability of the sediments to remove the elements from sea water and on dilution by non-hydrogenous materials. Turbidite and pelagic sediments differ in both respects. Slope sediments, from which turbidites arise, accumulate rapidly with higher organic content, whereas pelagites accumulate slowly with a more significant oxyhydroxide component. A fraction of the Pt in both pelagic and turbidite sediments is labile, indicating that Pt may be scavenged from sea water into both oxyhydroxide- and organic-rich sediments. In contrast, Ir mobility is observed clearly only in the reduction of oxic sediments. This difference between Ir and Pt may be accounted for by the ~50-fold lower Ir abundance in sea water and the inability of more rapidly accumulating slope sediments to sequester a significant amount of Ir (seawater concentrations: Ir < 8 femtomolar (fM) and Pt $\approx$ 300 fM; refs 17, 25, 26). The occurrence of mobile Re, on the other hand, is clearly controlled by the difference in sediment composition between the turbidites and pelagites. Re is strongly enriched in organic-rich sediments owing to reduction of the

ReO_4^- oxyanion, which is stable in sea water under oxic conditions. Oxic pelagic sediments do not scavenge a significant amount of Re from sea water, even with their slower accumulation rates and a relatively high seawater Re concentration (45,000 fM)^{17,25}.

This study demonstrates that variations in Ir concentrations (and changes in PGE ratios) in the geological record may be created by redox processes within the sediment. In particular, Fe–Mn oxyhydroxides contain a significant proportion of the Ir and Pt in oxic sediments and this fraction may be remobilized if oxides are reduced. Although the processes studied here cannot produce spikes of the magnitude observed at the K/T boundary, many smaller enrichments reported at other extinction horizons (Cenomanian/Turonian¹⁰, Frasnian/Famennian¹¹, Devonian/Carboniferous⁹) may be explained by the cycling of Ir with oxyhydroxide minerals; in these cases, Ir and Pt enrichments are accompanied by enhanced Fe, Mn, Co and Ni concentrations. The relatively modest elemental redistributions observed here are brief, non-steady-state diffusive phenomena. If these processes were sustained over longer times, or if advection of pore or ground waters were involved, greater diagenetic enrichments would result¹⁴. Other sedimentary components that might control Ir mobility (such as organic matter or sulphides) need to be more fully investigated in modern environments. □

Received 10 April; accepted 15 June 1992.

- Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. *Science* **208**, 1095–1108 (1980).
- Alvarez, W., Asaro, F., Michel, H. V. & Alvarez, L. W. *Science* **216**, 886–888 (1982).
- Xu, D.-Y. *et al. Nature* **314**, 154–156 (1985).
- Ganapathy, R. *Science* **216**, 885–886 (1982).
- Holser, W. T. *et al. Nature* **337**, 39–44 (1989).
- Hsu, K. J. *et al. Nature* **316**, 809–811 (1985).
- Kyte, F. T., Zhou, Z. & Wasson, J. T. *Nature* **292**, 417–420 (1981).
- Orth, C. J., Quintana, L. R., Gilmore, J. S., Grayson, R. C. Jr & Westergaard, E. H. *Geology* **14**, 986–990 (1986).
- Orth, C. J. *et al. Geology* **16**, 627–630 (1988).
- Orth, C. J. *et al. Geophys. Res. Lett.* **15**, 346–349 (1988).
- Playford, P. E., McLaren, D. J., Orth, C. J., Gilmore, J. S. & Goodfellow, W. D. *Science* **226**, 437–439 (1984).
- Wallace, M. W., Keays, R. R. & Gostin, V. A. *Geology* **19**, 551–554 (1991).
- Schmitz, B. *Geochim. cosmochim. Acta* **49**, 2361–2370 (1985).
- Wallace, M. W., Gostin, V. A. & Keays, R. R. *Geology* **18**, 132–135 (1990).
- Barker, J. & Anders, E. *Geochim. cosmochim. Acta* **32**, 627–645 (1968).
- Crocket, J. H. & Kuo, H. Y. *Geochim. cosmochim. Acta* **43**, 831–842 (1979).

- Crocket, J. H. & Kuo, H. Y. *Geochim. cosmochim. Acta* **43**, 831–842 (1979).
- Goldberg, E. D., Hodge, V., Kay, P., Stallard, M. & Koide, M. *Appl. Geochem.* **1**, 227–232 (1986).
- Koide, M. *et al. Appl. Geochem.* **1**, 705–714 (1986).
- Colley, S., Thomson, J., Wilson, T. & Higgs, N. *Geochim. cosmochim. Acta* **48**, 1223–1235 (1984).
- Thomson, J., Higgs, N. C. & Colley, S. *Mar. Geol.* **89**, 315–330 (1989).
- Thomson, J., Higgs, N. C., Croudace, I. W., Colley, S. & Hydes, D. J. *Geochim. cosmochim. Acta* (in the press).
- Wilson, T. R. S. *et al. Geochim. cosmochim. Acta* **49**, 811–822 (1985).
- Jarvis, I. & Higgs, N. in *Geology and Geochemistry of Abyssal Plains* (eds Weaver, P. P. E. & Thomson, J.) 179–213 (Geol. Soc. Spec. Publ., London, 1987).
- Colley, S., Thomson, J. & Toole, J. *Geochim. cosmochim. Acta* **53**, 1223–1234 (1989).
- Colodner, D. C. thesis, Massachusetts Inst. Technol.—Woods Hole Oceanogr. Inst. (1991).
- Jacinto, G. S. & van den Berg, C. M. G. *Nature* **338**, 332–334 (1989).
- Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363–2380 (1982).
- Kyte, F. T., Zhou, Z. & Wasson, J. T. *Nature* **288**, 651–656 (1980).
- Ganapathy, R. *Science* **209**, 921–923 (1980).

ACKNOWLEDGEMENTS. This work was supported by a NSF Graduate Fellowship, the Office of Naval Research, and a Woods Hole Oceanographic Institution Ocean Ventures Fund Fellowship.

Evidence for dipolar fields during the Cobb Mountain geomagnetic polarity reversals

Bradford M. Clement

Department of Geology, Florida International University, University Park, Miami, Florida 33199, USA

RECENT palaeomagnetic studies¹⁻⁴ of geomagnetic polarity reversals have presented conflicting interpretations of the behaviour of the Earth's magnetic field during reversals. Some have argued that the grouping of transitional virtual geomagnetic poles (VGPs) suggests the presence of large-scale, perhaps even dipolar, symmetries during some reversals¹⁻³, whereas others have claimed that the available data do not support the interpretation of simple transitional field geometries⁴. An earlier comparison⁵ of two North Atlantic sedimentary records of reversals bounding the Cobb Mountain subchron (1.1 Myr ago) with a volcanic record from Tahiti⁶ revealed striking similarities in the sequences of transitional VGPs, suggesting the presence of large-scale symmetries in these fields. Here I augment this comparison with two new records of the Cobb Mountain subchron from the western Pacific⁷, which greatly extend the geographical and temporal coverage. These records exhibit sequences of VGP positions that are very similar to those observed in the North Atlantic and Tahiti, providing evidence of dipolar transitional fields during the Cobb Mountain reversals.

The North Atlantic records of the Cobb Mountain subchron were obtained from deep-sea sediments cored at Deep Sea Drilling Project (DSDP) site 609 (49.86° N, 335.77° E)⁸ and Ocean Drilling Program (ODP) site 647 (53.32° N, 314.75° E)⁵. The Cobb Mountain event is defined in these records as a short (25,000-yr) interval of full normal polarity. Although the site 647 record contains less detail, the two records show striking similarities. The polarity transitions exhibit a marked anti-symmetry which is particularly well defined in the site 609 record (Fig. 1). The volcanic record of the Cobb Mountain event from Tahiti (17° S, 210.33° E)⁶ exhibits VGPs that fall near South Africa, off the southwest coast of Australia and in the central North Pacific, along the reverse-to-normal (R-N) reversal path from site 609 (Fig. 1).

Two additional records from the western Pacific greatly extend the geographical range of sites documenting the Cobb Mountain reversals. These records were obtained from ODP sites 767 (4.79° N, 123.50° E) and 768 (8.00° N, 121.22° E) in the Celebes and Sulu Seas, respectively⁷. The magnetization of these sediments was measured using detailed pass-through magnetometer measurements (0.5-cm sampling interval) and discrete samples (5-cm sampling interval)⁷. Although the pass-through measurements were closely spaced, the response functions for the magnetometer sensors average over up to 25 cm of stratigraphic section. Hsu *et al.*⁷ deconvolved the pass-through measurements by first eliminating frequencies higher than the instrument responses using a cosine taper window and then taking the inverse fast Fourier transform of the ratio of the power of the data to the instrument response.

The deconvolution seems to have recovered successfully some higher-frequency information from the western Pacific pass-through data, but the resulting records are still very smoothed by the measurement technique. The effects of smoothing palaeomagnetic records of polarity transitions have been well documented⁹⁻¹¹. Smoothing a sequence of directions tends to dampen sharp changes in direction and causes VGPs to fall along great-circle paths. Although the similarities between the western Pacific records and the original 609 record are apparent (Fig. 1), I artificially smoothed the 609 records by taking running unit-vector means of the 609 VGP paths to illustrate the effects

of smoothing. I calculated unit-vector means instead of total-vector means because of the uncertainty in interpreting the site 609 relative-intensity record as a measure of true palaeointensities. Averaging total vectors generally reduces the smoothing effect over intervals of low intensities, and therefore the averaging inherent in the pass-through measurements process may explain the lack of detail in the western Pacific records.

The Cobb Mountain record from site 767 bears a number of pronounced similarities to the 609 record, most notably in the R-N reversal, in which the VGPs loop over South Africa, return to the south geographical pole and then move northwards through the Pacific (Fig. 2). Within the northward path, a swing toward South Australia occurs, which is very similar to the swing

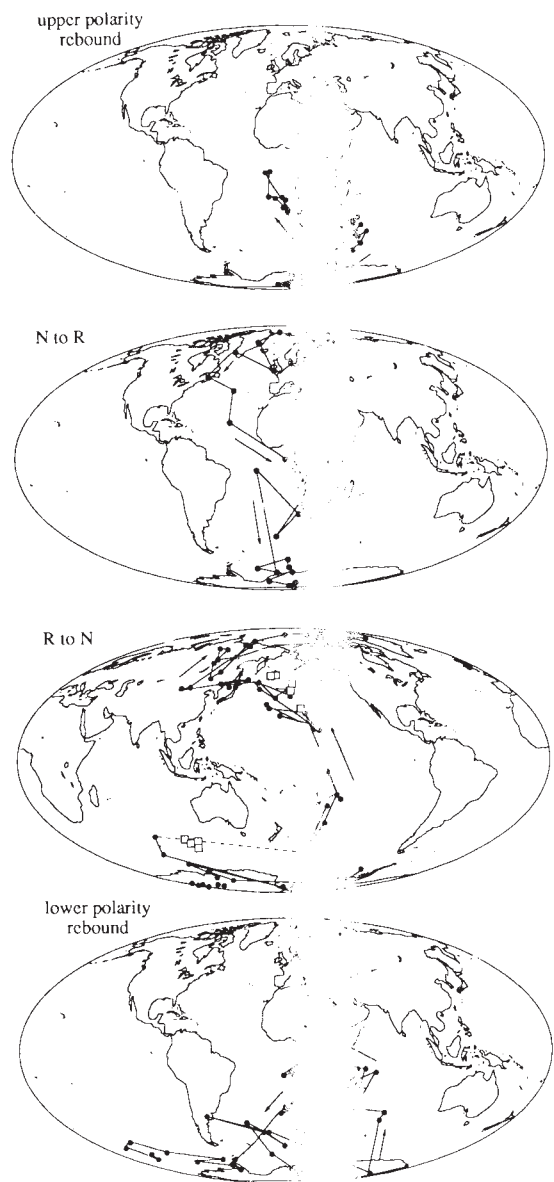


FIG. 1 Virtual geomagnetic pole (VGP) positions during the polarity transitions bounding the Cobb Mountain subchron as recorded in deep-sea sediments cored at DSDP site 609 in the North Atlantic. The VGP paths exhibit a distinct antisymmetry with nearly antipodal paths in the R-N and N-R reversals and with loops of opposite direction over South Africa occurring in both reversals. The transitional VGP clusters from the volcanic record of the Cobb Mountain⁶ (open squares) fall along the site 609 R-N reversal path, suggesting the presence of large-scale symmetries in this transitional field (see ref. 5 for a detailed discussion).

to positions off the southwest coast of Australia that occurs in the 609 record. The R-N VGP path from site 768 also exhibits important similarities to the 609 VGP path. Although the record appears to be more heavily smoothed, the VGPs swing toward south Africa and then progress northwards through the central Pacific.

In the N-R reversal path the VGPs in the site 767 record move southward through Africa in a path that falls close, but to the east of, the smoothed 609 N-R path (Fig. 3). The N-R VGP path from site 768 is the one of the three reversal records from the western Pacific that appears very different from the 609 record. In this record, the VGPs move southward through the Americas instead of through the Atlantic and Africa as do the 609 VGPs (Fig. 3). Hsu *et al.*⁷ interpreted the difference between the 767 and 768 N-R reversal paths as evidence that the transitional fields were characterized by very complex field geometries.

It is possible that the 768 N-R record differs because of effects associated with the pass-through measurement process. The 768 N-R discrete sample measurements did not record any intermediate declination values (unlike the other three western Pacific records); instead, the declinations have a sharp 180° change⁷. The pass-through measurements may have artificially introduced a trend by convolving a step function in declination values. To examine this possibility I changed the sense of the declination change (through westward rather than eastward directions) without changing the inclinations. The resulting VGP path (Fig. 3) is in much better agreement with the 609 record, suggesting that the discrepancy of 768 N-R record may indeed be related to the effects of the pass-through measurement and deconvolution process.

The five records discussed here show important similarities in both the positions and sequence of transitional VGPs in the

reversals bounding the Cobb Mountain subchron, particularly in the R-N reversal. These transitions were obtained from different types of palaeomagnetic records (lavas and deep-sea sediments) from widely separated locations, and therefore it is difficult to ascribe the similarities in these records to artefacts of the recording processes. The similarities are significant enough to suggest that, to within the resolution of the records, the transitional fields were predominantly dipolar.

The low intensities observed during the Cobb Mountain reversals indicate that the large-scale symmetries in these records occurred while the field was weak^{5,6,8}. The low intensities observed in many polarity transitions have been widely interpreted as evidence that the field reorganizes during reversals^{12,13}. This interpretation has served as the foundation for reversal models in which more complex field geometries dominate the transitional field¹⁴⁻¹⁶. Although the Cobb Mountain intensity data are consistent with a reorganization of the field, the directional data suggest that the reorganization resulted in fields characterized by large-scale symmetries.

Bogue and Merrill¹⁷ have suggested that a dipolar-like transitional field does not necessarily indicate organized field structure within the Earth's core. They point out that a white-noise spectrum at the core-mantle boundary may result in a surface field in which 50% of the total intensity is due to the dipole term. This is possible because the spherical-harmonic terms vary with distance from the centre of the Earth, with the shorter-wavelength terms falling off most quickly with distance. It should not be surprising, therefore, that a dipolar-like field may be

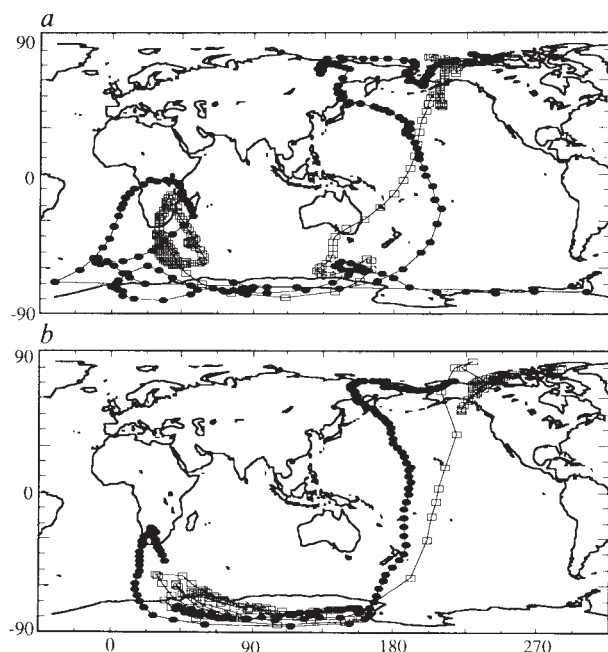


FIG. 2 *a*, The VGPs from the reverse-to-normal Cobb Mountain reversal as recorded at ODP site 767 (open symbols) in the Celebes Sea and the smoothed VGP path from DSDP site 609 (solid symbols) smoothed using a 5-cm window. *b*, The VGPs from site 768 (open symbols) in the Sulu Sea compared with the 609 VGP path smoothed using a 15-cm window. Although the pass-through records from sites 767 and 768 are greatly smoothed by the measurement process, both records exhibit a movement of VGPs toward positions over South Africa and then return to the south geographical pole before moving northward through the Pacific, in agreement with the sequence of VGP positions recorded at site 609.

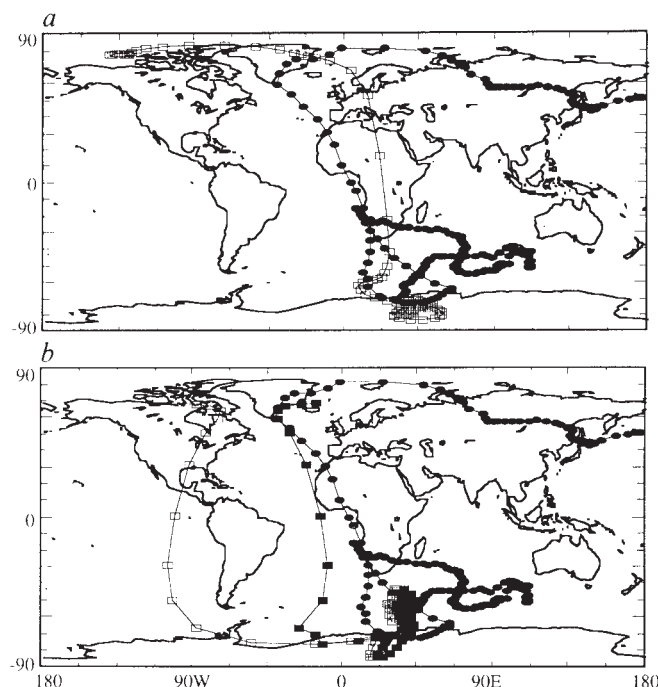


FIG. 3 *a*, The VGPs from the normal-to-reverse Cobb Mountain reversal as recorded at ODP site 767 (open symbols) in the Celebes and the smoothed (2.5-cm window) VGP path from DSDP site 609. The record from site 767 is very similar to the record from 609, with VGPs moving southwards through Africa. *b*, The VGP path from 768 (open symbols) in the Sulu Sea and the VGP path from DSDP site 609 (solid circles) smoothed using a 2.5-cm window. The VGP path from site 768 differs significantly from the 609 record, tracking southwards through the Americas rather than through Africa. The solid square symbols represent the site 768 path recalculated assuming that the pass-through measurements convolved a step function in declinations (see text). The similarity between this path and the site 609 path suggests that the discrepancy of 768 N-R path with the other records may be related to effects caused by the pass-through measurement and deconvolution process.

observed during some reversals, and this scenario may account for the apparently simple symmetries and the low intensities in the Cobb Mountain records. The recurrence of similar VGP positions in both the N-R and R-N reversals, such as the loops over South Africa, however, suggests something more systematic than the randomness implicit in Bogue and Merrill's white-noise spectrum. The recurrent VGP positions suggest either that the reversing field organizes itself as an intrinsic part of the reversal process or that factors external to the core, such as lateral variations in properties of the lowermost mantle, influence the reversing field.

The recurrence of similar geographical distributions of VGP paths during some reversals has been interpreted as evidence that some form of lateral heterogeneity in the lowermost mantle provides the geographical control. During several reversals, including the Matuyama-Brunhes, there seems to be a preference for the transitional VGPs to fall along two nearly antipodal longitudinal bands centred over the Americas and eastern Asia¹⁻³. Laj *et al.*³ suggested that there is a correlation between the preferred longitudinal bands and regions of fast seismic-wave propagation in the lower mantle¹⁸, which they interpreted as evidence that the mantle influences the reversing field.

The Cobb Mountain VGPs also tend to fall along antipodal tracks, but these tracks are centred over Africa and the central Pacific and are distinctly different from the preferred longitudinal grouping in other reversals. In fact, in the same maps used for the correlation of Laj *et al.*^{3,18}, the Cobb Mountain VGP paths correlate better with regions of slow seismic-wave propagation. Because these reversals occurred over time intervals too short for significant changes to have occurred in the lower mantle, the differences in these transitional fields are difficult to reconcile with the hypothesis of lower-mantle control. If the mantle does affect the field, the different transitional fields of these reversals implies the geodynamo responds differently to the same mantle conditions. The variations in the transitional fields of different reversals may ultimately provide some insight into the mechanisms through which the influence may occur^{19,20}.

The symmetries defined by antipodal VGP positions, such as those observed in the Matuyama-Brunhes, Cobb Mountain, Upper Olduvai and some other reversals¹⁻³, may reflect a fundamental character of some transitional fields. This is supported by models in which reversals result from interactions between dynamo symmetry families that are either symmetric or anti-symmetric about the Equator^{21,22}. The interaction of symmetry families intrinsic to the geodynamo with geographical influences involving the lower mantle may provide an explanation for the differences between the transitional fields of different reversals.

Regardless of whether the large-scale symmetries are superimposed on the field by the mantle or result from intrinsic properties of the geodynamo, it now seems that at least some transitional fields were characterized by relatively simple geometries. Whereas the available data for some reversals, such as the Matuyama-Brunhes, suggest large-scale, non-dipolar transitional fields², in the case of the Cobb Mountain reversals the transitional field geometries cannot be distinguished from dipolar fields. Any explanation of polarity reversal mechanisms or mantle interactions with the core must account for the presence of large-scale symmetries in these transitional fields. □

13. Fuller, M., Williams, I. & Hoffman, K. A. *Rev. Geophys.* **17**, 179-203 (1979).
14. Williams, I. & Fuller, M. *J. geophys. Res.* **86**, 11657-11665 (1981).
15. Constable, C. G. *J. geophys. Res.* **95**, 4587-4596 (1990).
16. Weeks, R. J., Fuller, M. & Williams, I. *J. geophys. Res.* **93**, 11613-11620 (1988).
17. Bogue, S. W. & Merrill, R. T. A. *Rev. Earth planet. Sci.* **20**, 181-219 (1992).
18. Dziewonski, A. M. & Woodhouse, J. H. *Science* **236**, 37-48 (1987).
19. Merrill, R. T. *et al. Phys. Earth planet. Inter.* **64**, 87-101 (1990).
20. Jeanloz, R. A. *Rev. Earth planet. Sci.* **18**, 347-386 (1990).
21. Hoffman, K. A. *Nature* **354**, 273-277 (1991).
22. Merrill, R. T. & McFadden, P. L. *J. geophys. Res.* **93**, 11589-11597 (1988).

ACKNOWLEDGEMENTS. I thank V. Hsu for providing the data from sites 767 and 768. This work was supported by the NSF.

New 'phantom' dinoflagellate is the causative agent of major estuarine fish kills

JoAnn M. Burkholder, Edward J. Noga*,
Cecil H. Hobbs & Howard B. Glasgow Jr

Department of Botany, Box 7612, North Carolina State University, Raleigh, North Carolina 27695-7612, USA

* Department of Companion Animal and Special Species Medicine, North Carolina State University, 4700 Hillsborough Street, Raleigh, North Carolina 27606, USA

A WORLDWIDE increase in toxic phytoplankton blooms over the past 20 years^{1,2} has coincided with increasing reports of fish diseases and deaths of unknown cause³. Among estuaries that have been repeatedly associated with unexplained fish kills on the western Atlantic Coast are the Pamlico and Neuse Estuaries of the southeastern United States⁴. Here we describe a new toxic dinoflagellate with 'phantom-like' behaviour that has been identified as the causative agent of a significant portion of the fish kills in these estuaries, and which may also be active in other geographic regions. The alga requires live finfish or their fresh excreta for encystment and release of a potent toxin. Low cell densities cause neurotoxic signs and fish death, followed by rapid algal encystment and dormancy unless live fish are added. This dinoflagellate was abundant in the water during major fish kills in local estuaries, but only while fish were dying; within several hours of death where carcasses were still present, the flagellated vegetative algal population had encysted and settled back to the sediments. Isolates from each event were highly lethal to finfish and shellfish in laboratory bioassays. Given its broad temperature and salinity tolerance, and its stimulation by phosphate enrichment, this toxic phytoplankton may be a widespread but undetected source of fish mortality in nutrient-enriched estuaries.

The alga, which represents a new family, genus and species within the order Dinamoebales (K. Steidinger, personal communication), was inadvertently discovered by fish pathologists⁵ who observed sudden death of cultured tilapia (*Oreochromis aureus* and *O. mossambica*) several days after their exposure to freshly collected water from the Pamlico River. A small dinoflagellate increased in abundance before fish death, followed by a sharp decline unless additional live fish were introduced. Within 1-2 h of fish death, the sudden algal decline occurred because the cells produced resting cysts or non-toxic amoeboid stages and settled to the bottom of culture vessels (J.M.B., unpublished results).

The behaviour of the culture contaminant provided clues about its potential activity in estuarine habitat. We suspected that if low cell densities were lethal to fish in the wild, as in culture, then the alga would probably not be detected by routine monitoring efforts but might be found while a kill was in progress. We observed the dinoflagellate swarming in Pamlico water collected during a kill of nearly one million Atlantic menhaden (*Brevoortia tyrannus*; Table 1). Less than one day after the kill, few toxic vegetative cells remained suspended in the water

Received 16 March; accepted 8 June 1992.

1. Tric, E. *et al. Phys. Earth planet. Inter.* **63**, 319-336 (1991).
2. Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48-58 (1991).
3. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).
4. Valet, J.-P., Tucholke, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400-407 (1992).
5. Clement, B. M. & Martinson, D. G. *J. geophys. Res.* **97**, 1735-1752 (1992).
6. Chauvin, A., Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2727-2752 (1990).
7. Hsu, V., Merrill, D. L. & Shibuya, H. *Geophys. Res. Lett.* **17**, 2069-2072 (1990).
8. Clement, B. M. & Kent, D. V. *Earth planet. Sci. Lett.* **81**, 253-264 (1987).
9. Watkins, N. D. *Geophys. J. R. astr. Soc.* **17**, 121-149 (1969).
10. Hoffmann, K. A. & Slade, S. B. *Geophys. Res. Lett.* **13**, 483-486 (1986).
11. Rochette, P. *Earth planet. Sci. Lett.* **98**, 33-39 (1990).
12. Hoffman, K. A. *Science* **196**, 1329-1332 (1977).

although many fish carcasses still were present. In standard bioassays^{6,7} maximum algal growth coincided with fish death, followed by rapid encystment (Fig. 1). This toxic dinoflagellate was found in abundance during several other major fish kills in the Pamlico and Neuse estuaries and in local aquaculture facilities (Table 1). Isolates from these events have been confirmed in laboratory bioassays as lethal to 11 species of finfish including commercially valuable striped bass (*Morone saxatilis*), southern flounder (*Paralichthys lethostigma*), menhaden and eel (*Anguilla rostrata*).

The algal isolates all exhibit similar behaviour. Flagellated, photosynthetic toxic vegetative cells excyst after live fish or their fresh excreta are added to aquaria cultures under low light (Fig. 2a, b). The lag period for excystment ranges from minutes to days and increases with dormancy period or cyst age. The cells appear athecate but have thin cellulosic deposits (thecal plates, Fig. 2c)⁸ beneath outer membranes that obscure the plate arrangement required for formal speciation⁹. This dinoflagellate completes its sexual cycle while killing fish. Vegetative cells produce poorly pigmented, anisogamous gametes; the male is smaller, and its extended longitudinal flagellum is five- to sixfold longer than the cell (Fig. 2d). The vegetative cells can also form large amoeboid stages (pigmented or colourless, 80–250 μm long with extended pseudopodia) with unknown role in the life cycle.

TABLE 1 Fish kills associated with the new toxic dinoflagellate, documented in North Carolina during 1991–92

Date	Location	Temperature (°C)	Salinity (‰)	Fish killed	Alga (cells ml^{-1})
1991					
May	Pamlico River	24	3	Menhaden	1,300
June	Pamlico River	29	10	Menhaden, others	1,100
Aug.	Pamlico River	31	12	Menhaden	600
Aug.	Pamlico River	30	8	Flounder walk	800
Sept.–Oct.	Neuse River	26	11	Menhaden, Blue crabs	1,200
Dec.	Taylor's Creek	15	30	Fish walk (flounder, eel, mullet, others)	~35,000
1992					
Jan.	Aquaculture	9	0	Striped bass	
Feb.	NC maritime museum (Newport River)	21	25	Fish spp.	
Feb.	National Marine Fisheries Service (Newport River)	15–21	25	Menhaden, others	

Locations on the Pamlico River followed a salinity gradient from 3 to 12‰ over an 8-km distance. Low dissolved oxygen (DO , 2.8 mg l^{-1}) was measured from the bottom water in the vicinity of the kill during June 1991. On all other dates, DO was $\geq 4.5 \text{ mg l}^{-1}$ throughout the water column (R. Carpenter, K. Lynch and K. Miller, NC Dept. of Environmental Health and Natural Resources, personal communication). In each case from 1991, we observed abundant pigmented, flagellated vegetative cells in water samples taken during the kill. (Note, 'Walk' refers to cases in which stressed fish attempted to leave the water and beach before they died.) Water from fish kills in 1992 was collected within 1–4 days after fish death; hence, it was not possible to obtain data for cell abundances while the kills were in progress. Samples from a kill at a freshwater aquaculture facility in Aurora, NC contained ~40 vegetative cells ml^{-1} as well as cysts. Water from both kills in February 1992 (original source, the Newport River, 11 °C, 25‰) yielded toxic vegetative cells as well as gametes, planozygotes, amoeboid stages and cysts. We confirmed virtually identical morphology of these with previous isolates using scanning electron microscopy, and verified toxicity of each population using aquarium bioassays with tilapia and striped bass as test species.

Gamete fusion results in planozygote formation followed by production of additional toxic vegetative cells if live fish are present. Without live fish, the planozygotes lose their flagella and form thick-walled cysts but remaining gametes continue to multiply especially under phosphate enrichment (batch cultures enriched with 50–400 $\mu\text{g PO}_4^{3-} \text{ l}^{-1}$ yielded significantly more gametes after 6 days than cultures with $\text{PO}_4^{3-} \leq 10 \mu\text{g l}^{-1}$; Student's t test, $P < 0.01$)¹⁰. Such stimulatory effects have not been observed with either nitrate or ammonium. A ciliated protozoan, *Stylonichia* sp.¹¹, is common in local estuarine waters and consumes both toxic vegetative cells and cysts. But a pigmented, toxic amoeboid stage of the alga, in turn, attacks the protozoan predator.

Each flagellated vegetative algal cell forms a peduncle or pseudopodium shortly after excystment^{12,13}. The peduncle becomes fully extended during toxic activity (Fig. 3a), particularly at the optimal salinity of 15‰ (among tested salinities 5, 10, 15, 25 and 35‰; significantly higher vegetative cell production at 15‰, Student's t test, $P < 0.01$)¹⁰. The lethal agent is an excreted neurotoxin (under analysis by D. Baden, personal communication). Water from which cells had been removed by gentle drop-filtration (0.22 μm -pore filters) induces neurotoxic signs by fish including sudden sporadic movement, disorientation, lethargy and apparent suffocation followed by death. The alga has not been observed to attack fish directly. It rapidly increases its swimming velocity to reach flecks of sloughed tissue from dying fish, however, using its peduncle to attach to and digest the tissue debris. Within several hours of fish death, toxic vegetative cells encyst in the presence or absence of fish carcasses (Fig. 3b–d). The cysts can be destroyed by dilute bleach. But after treatment with concentrated sulphuric acid or ammonium hydroxide, 35 days of desiccation, or nearly 2 years of dormancy, small percentages of the cysts have yielded viable toxic cells when placed in saline water (10–15‰) with live fish (J.M.B., unpublished results).

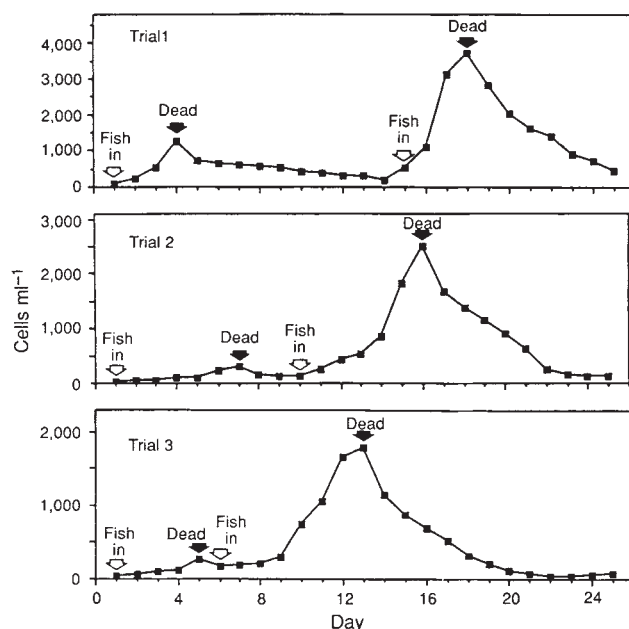
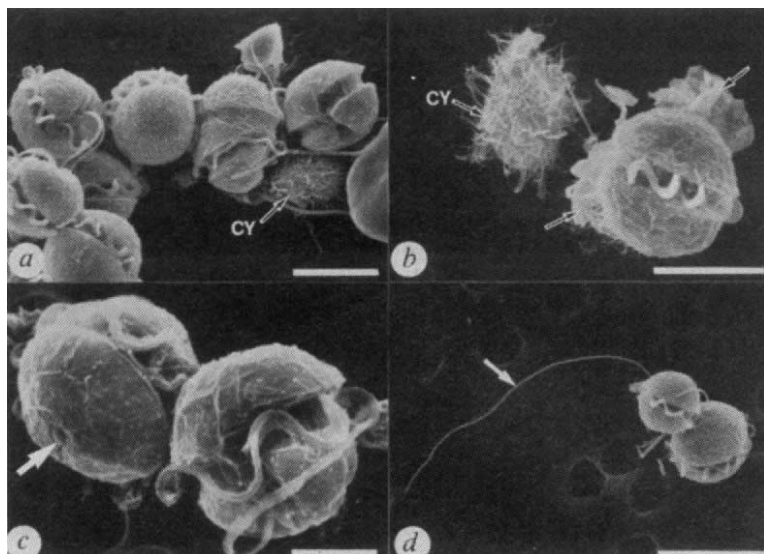


FIG. 1 Response of the dinoflagellate (toxic flagellated, vegetative stage) to *Oreochromis aureus* (length 3 cm, approximate age 40 days) in repeat-trial experiments. The trials were conducted as batch-culture aquarium bioassays (18 °C, 10‰, 30 $\mu\text{Einste m}^{-2} \text{ s}^{-1}$) which varied in the time interval between death of the first fish and addition of a second live fish (fish length 5 cm). The toxic vegetative stage attained greatest abundance just before fish death, followed by rapid decrease in abundance as the cells encysted or formed non-toxic, colourless amoebae and settled out. As we reduced the time interval without live fish, the second fish died more rapidly.

FIG. 2 Scanning electron micrographs of the toxic dinoflagellate. *a, b, c*, Several hours after a live tilapia was added to aquarium cultures (with environmental characteristics described in Fig. 1) after a 2-day period without fish. Most cysts (cy) had already produced flagellated vegetative cells (*a*, scale bar 10 μm ; *b*, arrows indicate remains of cyst wall; scale bar, 8 μm); *c*, Cell showing the outline of plates or 'armour' (for example, arrow indicating an apical plate) beneath surficial membranes (scale bar, 5 μm); *d*, Male and female gametes (left and right, respectively). The presumed gametes were abundant ~ 1 day after addition of a live fish, when the fish began to exhibit neurotoxic signs of lethargy, inability to maintain balance, and apparent respiratory distress. In viewing live samples under light microscopy, we have observed such cells in fusion; note the extended length of the longitudinal flagellum on the male (arrow)(scale bar, 15 μm). Photos by C. Hobbs.

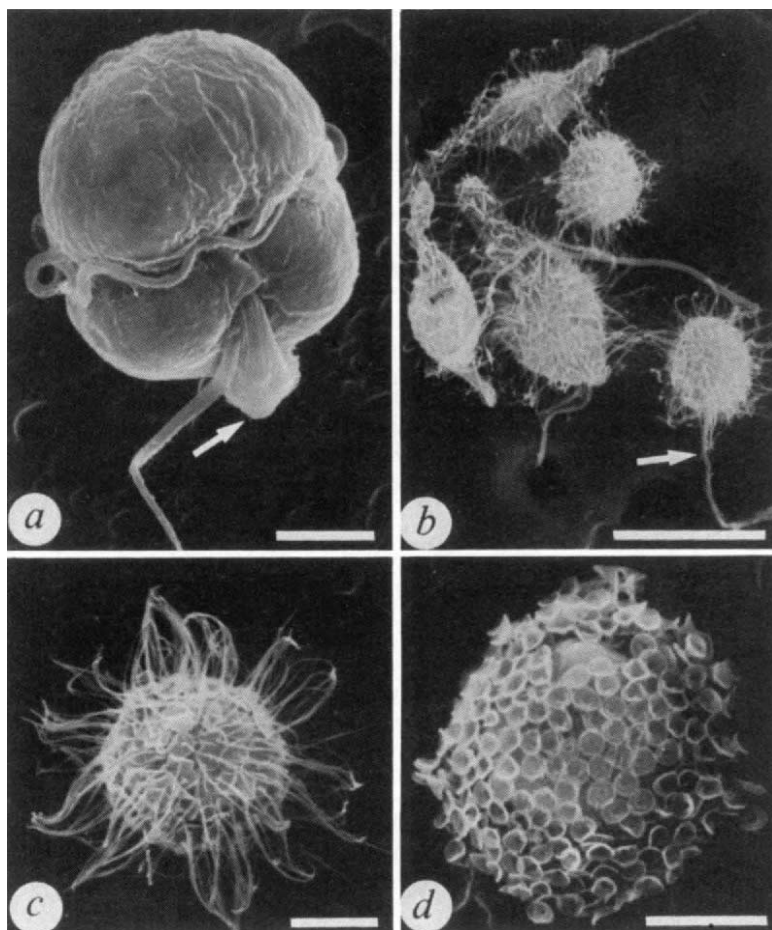


Laboratory bioassays and field collections have confirmed that this dinoflagellate is lethal to finfish across a temperature gradient from 4 to 28 $^{\circ}\text{C}$ ¹⁴⁻¹⁷, and across a salinity range of 2-35‰. It has also killed hybrid striped bass *Morone saxatilis* $\times$ *M. chrysops* in fresh water (0‰ salinity, in aquaculture facilities and in bioassays) with moderate concentrations of divalent cations (Table 1). Both native and exotic fishes are affected, suggesting stimulation by amino acids or other common, labile substance in fish excreta^{18,19} (Table 1). Shellfish such as blue

crabs (*Callinectes sapidus*, carapace width 8-10 cm) and bay scallops (*Aequipecten irradians*, shell width 4-5 cm) do not stimulate excystment or toxic activity. These shellfish remained viable for a 9-day test period while filtering low concentrations of the dinoflagellate (~ 50 cells ml^{-1}), although the scallop closing reflex slowed perceptibly. When placed in aquaria with dying finfish, however, blue crabs were killed within hours to several days, and scallops died within minutes.

Long-term data sets from Europe, Asia, and North America

FIG. 3 Scanning electron micrographs of the dinoflagellate from aquarium cultures. *a*, While tilapia were dying. At its optimal salinity of 15‰, the flagellated vegetative cells assume a swollen appearance and the peduncle (arrow) becomes fully extended for use in saprotrophic feeding on bits of fish tissue (scale bar, 3 μm); *b*, Two hours after the tilapia died and were not replaced with live fish. Most vegetative cells had begun to form cysts; note that the longitudinal flagellum was still attached to one of the cells (arrow)(scale bar, 10 μm); *c*, One day after fish death, many completed cysts presumed to contain dinoflagellate cells were present. Note that the outer covering consists of scales with long bristles (scale bar, 3 μm); *d*, Nearly 1 month after fish death, the outer scales on most cysts had lost the bristle extensions (scale bar, 5 μm).



strongly correlate toxic phytoplankton blooms with increasing nutrient enrichment²⁰. The Pamlico and Neuse Estuaries receive high anthropogenic loading of phosphorus and nitrogen^{21,22}. Over the past century, cultural eutrophication has probably shifted the habitat to more favourable conditions for algal growth and toxic activity. Unlike other toxic phytoplankton, however, this dinoflagellate is ephemeral in the water column. The lethal flagellated vegetative cells move from the sediment surface to the water in response to lingering finfish which become

lethargic from the toxin. The alga kills the organisms which stimulate it and then rapidly descends. Increasing reports worldwide describe unexplained kills involving fish that die quickly after exhibiting neurotoxic signs^{23,24}. It is unlikely that this eurythermal, euryhaline dinoflagellate occurs only in these estuaries. We predict that with well timed sampling, this alga will be discovered at the scene of many fish kills in shallow, turbid, eutrophic coastal waters extending to geographic regions well beyond the Pamlico and the Neuse. □

Received 6 April; accepted 5 June 1992.

1. Hallegraeff, G. M., Steffensen, D. A. & Wetherbee, R. J. *Plankt. Res.* **10**, 533–541 (1988).
2. Robineau, B., Gagne, J. A., Fortier, L. & Cembella, A. D. *Mar. Biol.* **108**, 293–301 (1991).
3. Shumway, S. E. *J. World Aquaculture Soc.* **21**, 65–104 (1990).
4. Miller, K. H. et al. *Pamlico Environmental Response Team Report* (North Carolina Department of Environment, Health and Natural Resources, Wilmington, 1990).
5. Smith, S., Noga, E. J. & Bullis, R. A. *Proc. 3rd Int. Colloquium Path. Mar. Aquaculture* 167–168 (1988).
6. Ryther, J. H. *Ecology* **35**, 522–533 (1954).
7. Roberts, R. J., Bullock, A. M., Turner, M. K., Jones, K. & Tett, P. *J. mar. Biol. Ass. UK* **63**, 741–743 (1983).
8. Taylor, F. J. R. in *The Biology of Dinoflagellates* (ed. Taylor, F. J. R.) 24–92 (Blackwell, Boston, 1987).
9. Taylor, F. J. R. in *The Biology of Dinoflagellates* (ed. Taylor, F. J. R.) 723–731 (Blackwell, Boston, 1987).
10. Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (Freeman, San Francisco, 1981).
11. Thorp, J. H. & Covich, A. P. *Ecology and Classification of North American Freshwater Invertebrates*, 77 (Academic, New York, 1991).
12. Spero, H. J. *J. Phycol.* **18**, 357–360 (1982).
13. Gaines, G. & Elbrachter, M. in *The Biology of Dinoflagellates* (ed. Taylor, F. J. R.) 224–268 (Blackwell, Boston, 1987).
14. North Carolina Department of Environment, Health and Natural Resources. *1990 Algal Bloom Reports* (North Carolina Department of Environmental Management, Raleigh, 1990).
15. North Carolina Department of Natural Resources and Community Development. *1986 Algal Bloom Reports* (North Carolina Division of Environmental Management, Water Quality Section, Raleigh, 1987).

16. North Carolina Department of Natural Resources and Community Development. *1987 Algal Bloom Reports* (North Carolina Division of Environmental Management, Water Quality Section, Raleigh, 1988).
17. North Carolina Department of Natural Resources and Community Development. *1988 Algal Bloom Reports* (North Carolina Division of Environmental Management, Water Quality Section, Raleigh, 1989).
18. Palenik, B., Kieber, D. J. & Morel, F. M. M. *Biol. Oceanogr.* **6**, 347–354 (1988).
19. Lagler, K. F., Bardach, J. E. & Miller, R. R. *Ichthyology* 268–272 (New York, 1962).
20. Smayda, T. J. in *Novel Phytoplankton Blooms* (ed. Cosper, E. M. et al.) 449–484 (Springer, New York, 1989).
21. Stanley, D. H. in *Proc. Symp. Coastal Water Resources* 155–164 (American Water Works Association, Wilmington, 1988).
22. Rudek, J., Paerl, H. W., Mallin, M. A. & Bates, P. W. *Mar. Ecol. Prog. Ser.* **75**, 133–142 (1991).
23. White, A. W. *Mar. Biol.* **65**, 255–260 (1981).
24. White, A. W. in *Proc. Int. Conf. Impact of Toxic Algae on Mariculture* 9–14 (Aqua-Nor 1987 Int. Fish Farming Exhib., Trondheim 1988).

ACKNOWLEDGEMENTS. We thank K. Lynch, K. Miller and R. Carpenter (NC Dept. of Environment, Health and Natural Resources), and N. McNeill and W. Hettler (National Marine Fisheries Service) for their help in obtaining water samples and V. Coleman, J. Compton, C. Harrington, and C. Johnson for laboratory assistance. This work was supported by the Albemarle-Pamlico Estuarine Study through the UNC Water Resources Research Institute, and the Department of Botany and the College of Veterinary Medicine at North Carolina State University.

Physiological and behavioural thermoregulation in bigeye tuna (*Thunnus obesus*)

Kim N. Holland*, Richard W. Brill†, Randolph K. C. Chang†, John R. Sibert‡ & David A. Fournier‡

* Hawaii Institute of Marine Biology, PO Box 1346, Coconut Island, Kaneohe, Hawaii 96744, USA

† National Marine Fisheries Service, Honolulu Laboratory, 2570 Dole Street, Honolulu, Hawaii 96822, USA

‡ Otter Research Ltd, PO Box 265, Nanaimo, British Columbia, Canada

TUNA are unique among teleost fishes in being thermoconserving. Vascular counter-current heat exchangers maintain body temperatures above ambient water temperature, thereby improving locomotor muscle efficiency, especially at burst speeds and when pursuing prey below the thermocline^{1–6}. Because tuna also occasionally swim rapidly in warm surface waters, it has been hypothesized that tuna thermoregulate to accommodate changing activity levels or ambient temperatures⁷. But previous field experiments have been unable to demonstrate definitively short-latency, mammalian-type physiological thermoregulation^{8,9}. Here we show using telemetered data that free-ranging bigeye tuna (*Thunnus obesus*) can rapidly alter whole-body thermal conductivity by two orders of magnitude. The heat exchangers are disengaged to allow rapid warming as the tuna ascend from cold water into warmer surface waters, and are reactivated to conserve heat when they return into the depths. Combining physiological and behavioural thermoregulation expands the foraging space of bigeye tuna into otherwise prohibitively cold, deep water.

In all fish, oxygenated arterial blood leaves the gills essentially equilibrated to ambient water temperature (T_a). But in tunas, arterial blood passes through dense retia of interdigitated veins carrying warm blood away from the swimming muscles. The whole-body heat-transfer coefficient (k) is thereby reduced and heat is returned into the tissues^{1–3}. Although elevated body

temperature (T_b) improves locomotor muscle power for these obligate swimmers when they are at high speeds or below the thermocline^{4,5}, a permanently engaged heat retention system could be counterproductive for tuna swimming rapidly (fleeing, or chasing prey) in warm surface waters, because T_b may rise to injurious levels. Thus, it has been postulated that tuna can modify thermoconservation efficiency to fit the immediate demands of activity level and thermal environment⁷.

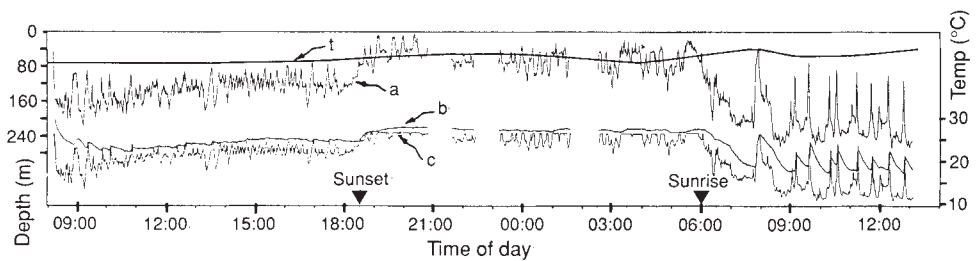
Studies with captive tuna have been suggestive of an ability to regulate k by a few per cent (probably through altered cardiac output which changes the dwell time of blood in the retia)^{7,10}, but data from field experiments with bluefin tuna (*T. thynnus*), are ambiguous^{8,9}, and a transmitter placed in a bigeye tuna stomach failed to demonstrate thermoregulation⁸. Hysteresis in the warming and cooling rates of swordfish, *Xiphias gladius*, occurs over periods of hours¹¹ and is associated with irregular diving patterns, making it difficult to separate the effects of changes in activity level, cardiac output and thermal inertia imparted by simple heat exchangers^{3,11–13}. Tuna, however, experience very rapid changes in T_a , which would require fast thermoregulatory adjustments. Thus, the question remained: can tuna perform physiological thermoregulation, or are the retia simply passive devices sensitive to blood flow rate?

The behaviour of subadult bigeye tuna in Hawaiian waters is ideal for testing for the occurrence of physiological thermoregulation. During daytime, they select temperatures between 14 and 17 °C (at depths around 250 m), from which they make regular periodic, rapid vertical excursions up into warmer waters¹⁴. During these excursions, the fish experience first increasing and then decreasing T_a , changing at up to 12 °C min⁻¹. The comparatively small size of these tuna (65 to 80 cm; 7.0 to 12 kg), reduces the influence of simple thermal inertia on T_b .

In open-ocean tracking experiments, we telemetered swimming depth and T_b data from two bigeye tuna, one of which exhibited typical bigeye distribution and behaviour (Fig. 1). The radically different rates of body warming and cooling observed during the vertical excursions were modelled with a numerical parameter estimation procedure¹⁵ which minimizes the squared differences between observed and estimated T_b .

Tuna body temperature is a function of heat exchange with

FIG. 1 Simultaneous swimming depth (a; left axis), T_b (b; right axis) and T_a (c; right axis) of bigeye tuna number 2 (length, 79 cm; weight ~ 11.7 kg). Acoustic telemetry techniques¹⁸ were used to monitor simultaneously swimming depth and T_b of a free-ranging bigeye tuna for 29 h. T_b was measured by a thermister in the tip of a 6.0×0.25 cm hypodermic needle inserted from the dorsal surface. Thermister probe length was pre-set to reach swimming muscles 2 to 3 cm dorso-lateral to the spine; angling the probe caudally prevented it from moving during swimming (as confirmed with captive fish). T_a was derived by interpolating between sequential expendable bathythermograph deployments and knowing the depth of the fish. T_b and depth data were recorded on tape on the tracking vessel. This fish demonstrated the typical diel pattern of swimming closer to the surface at night and, on the second



morning, the deep distribution and vertical excursions regularly observed in previous bigeye tracks¹⁴. T_b were buffered against changes in T_a resulting from brief movements up and down through the water column. Solid horizontal line (t) represents the interface between the bottom of the surface mixed layer and the top of the thermocline.

the environment and internal (metabolic) heat production ($\dot{T}_m$). In this case, the symmetry of the rates of ascent and descent, combined with constant horizontal speed over ground, indicated that locomotor output, and hence $\dot{T}_m$, was constant during the observed periods of warming and cooling. Heat loss (or gain) is proportional to the difference between body temperature and the surrounding water:

$$\frac{dT_b}{dt} = k(T_a - T_b) + \dot{T}_m \quad (1)$$

The estimated values of k for warming and cooling which

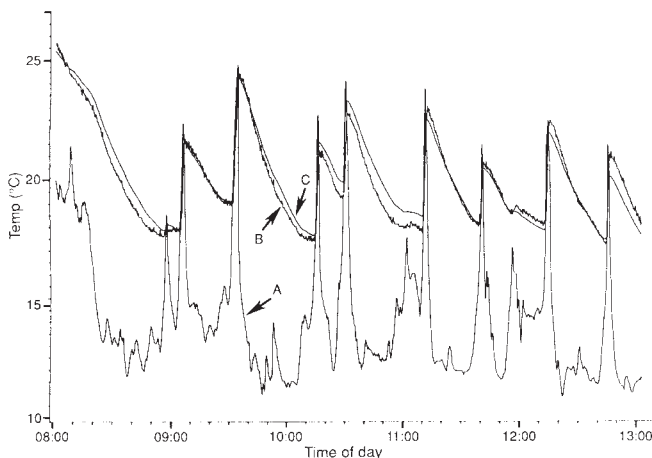


FIG. 2 Relationship between ambient temperature (A), observed body temperature (B), and estimated body temperature (C) for fish number 2, day 2. Using equation (1), and a switching function (equation (2)), automatic differentiation^(15,19) was used to fit two possible values of k to the observed data from fish 2

$$k = \begin{cases} k_1, & \text{if } T_a - T_b \leq \Delta T_{\text{crit}} \\ k_2, & \text{if } T_a - T_b > \Delta T_{\text{crit}} \end{cases} \quad (2)$$

where k_1 and k_2 are different values for whole-body heat-transfer coefficient such that $k_1 < k_2$, and $k = k_1$ when the fish is in heat-retention mode and $k = k_2$ when in heat-absorption mode. ΔT_{crit} is some threshold value for the difference between T_b and T_a , such that the heat exchanger is engaged when the fish is in water that is ΔT_{crit} cooler than T_b and disengaged when in water that is ΔT_{crit} degrees warmer than T_b . The model parameters were estimated in a stepwise fashion; first $\dot{T}_m$, then k_1 and k_2 and finally ΔT_{crit} . Improvement in fit at each step was tested using a simple F test on the decrease in the residual sum of squares. The T_b curves resulting from the modelled k_1 and k_2 values closely match the rapid warming and slow cooling observed during the vertical excursions. These excursions were consistently initiated when T_b declined to $\sim 17.5^\circ\text{C}$, possibly indicating a minimum tolerable T_b , even though still 6°C above T_a . The estimated value for $\dot{T}_m$ was $1.12 \times 10^{-4} \text{ }^\circ\text{C s}^{-1}$ and estimated value of ΔT_{crit} was 0.03°C , within the range of temperature detection thresholds observed experimentally with captive fish^{20,21}.

best fit the observed data (Fig. 2) are: $k_1 = 5.22 \times 10^{-4}$ (cooling), and $k_2 = 4.01 \times 10^{-2}$ (warming). These values indicate that the fish was rapidly varying the rate of instantaneous heat transfer by over two orders of magnitude. Examination of the onset of changes in T_b indicates that the conductivity switches were implemented in about 45 s.

The tunas have progressively evolved away from the typical teleost vasculature; blood supplying the swimming muscles is delivered and retrieved less through central blood vessels running through the haemal arch, and more through sub-dermal retia and cutaneous arteries and veins^{1-3,5}. In the comparatively primitive skipjack tuna (*Katsuwonus pelamis*), there are both central and lateral heat exchangers, whereas the most highly evolved (and most temperate in distribution) bluefin tuna no longer has a complete central blood supply to the musculature, but a complex lateral rete system. This progressive lateralization of blood supply may be a response to the radiation of tuna away from tropical environments; supplying the muscles entirely through lateral heat exchangers most effectively insulates the fish and limits the flow of cold arterial blood into the core⁵.

Bigeye tuna are intermediate in both their vascular design and distribution; they have well developed lateral heat exchangers, and retain a modest central blood supply, but it does not pass through a countercurrent rete^{2,5}. Routing arterial blood through the lateral rete system would put the tuna into a thermoconserving mode (k_1); passing the blood through the central system would result in an essentially poikilothermic fish (k_2). The latency of onset of warming is consistent with adrenergic neural control of vascular dilation¹⁶; in skipjack tuna at least, heat exchanger arteries have smooth muscle in their walls¹⁷.

Although inhabiting semitropical latitudes, bigeye tuna have expanded their range by invading the cooler depths below the surface mixed layer and evolving a behavioural repertoire and circulatory system that take advantage of the vertical proximity of warm water. Their vascular anatomy provides a capacity for short-latency physiological thermoregulation which, when coupled with behavioural thermoregulation expressed as the brief vertical excursions, allows them to forage in deep waters and still maintain the advantages of high muscle temperatures derived from their tropical lineage. □

Received 13 April; accepted 9 June 1992.

- Graham, J. B. & Diener, D. R. in *The Physiological Ecology of Tunas* (eds Sharp, G. D. & Dizon, A. E.) 113-134 (Academic, New York, 1978).
- Carey, F. G. *Sci. Am.* **228**, 36-44 (1973).
- Graham, J. B. in *Fish Biomechanics* (eds Webb, P. W. & Weihs, D.) 248-279 (Praeger, New York, 1983).
- Neill, W. H., Chang, R. K. C. & Dizon, A. E. *Environ. Biol. Fish.* **1**, 61-80 (1976).
- Graham, J. B. *Fish. Bull.* **73**, 219-229 (1975).
- Stevens, E. D. & Neill, W. H. in *Fish Physiology VII* (eds Hoar, W. S. & Randall, D. J.) 315-359 (Academic, New York, 1978).
- Dizon, A. E. & Brill, R. W. *Am. Zool.* **19**, 249-265 (1979).
- Carey, F. G. & Lawson, K. D. *Comp. Biochem. Physiol.* **44**, 375-392 (1973).

9. Neill, W. S. & Stevens, E. D. *Science* **184**, 1008–1010 (1974).
10. Graham, J. B. & Dickson, K. A. *Physiol. Zool.* **54**, 470–486 (1981).
11. Carey, F. G. in *Planning the Future of Billfishes* (ed. Stroud, R. H.) 103–122 (National Coalition Marine Conservation, Savannah, Georgia, 1990).
12. Carey, F. G. & Scharold, J. V. *Mar. Biol.* **106**, 329–342 (1990).
13. Carey, F. G. & Gibson, Q. H. *Physiol. Zool.* **60**, 138–148 (1987).
14. Holland, K. N., Brill, R. W. & Chang, R. K. C. *Fish. Bull.* **88**, 493–507 (1990).
15. *AUTODIFF: a C++ Array Language Extension with Automatic Differentiation for Use in Non-linear Modelling and Statistics* (Otter Research Ltd, Nanaimo, Canada, 1991).
16. Axelsson, M. & Nilsson, S. *J. exp. Biol.* **126**, 225–233 (1986).
17. Stevens, E. D., Lam, H. M. & Kendall, J. *J. exp. Biol.* **61**, 145–153 (1974).
18. Holland, K. N., Brill, R. W., Chang, R. K. C. & Yost, R. *Mar. Fish. Rev.* **47**, 26–32 (1985).
19. Greiwank, A. & Corliss, G. F. (eds) *Automatic Differentiation of Algorithms: Theory, Practice and Application* (SIAM, Philadelphia, 1991).
20. Dizon, A. E., Byles, T. C. & Stevens, E. D. *J. therm. Biol.* **1**, 185–187 (1976).
21. Steffel, S., Dizon, A. E., Magnuson, J. J. & Neill, W. H. *Trans. Am. Fish. Soc.* **105**, 588–591 (1976).

ACKNOWLEDGEMENTS. We thank E. D. Stevens and G. W. Boelhart for reviews of the manuscript and the Sea Grant College of the University of Hawaii and the National Marine Fisheries Service, Honolulu Laboratory, NOAA for their support of this research.

Neural correlates of perceptual motion coherence

Gene R. Stoner & Thomas D. Albright

Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

THE motions of overlapping contours in a visual scene may arise from the physical motion(s) of either a single or multiple surface(s). A central problem facing the visual motion system is that of assigning the most likely interpretation. The rules underlying this perceptual decision can be explored using a visual stimulus formed by superimposing two moving gratings. The resultant percept is either that of a single coherently moving 'plaid pattern' (coherent motion) or of the two component gratings sliding noncoherently across one another (noncoherent motion)^{1,2}. When plaid patterns are configured to mimic one transparent grating overlying another, the percept of noncoherent motion dominates³. We now report that neurons in the visual cortex of rhesus monkeys exhibit changes in direction tuning that parallel this perceptual phenomenon: sensitivity to the motions of the component gratings is enhanced under conditions that favour the perception of noncoherent motion. These results challenge models of cortical visual processing that fail to take into account the contribution of figural image segmentation cues to the analysis of visual motion.

The perceptual phenomenon of motion coherence may be mediated by a subset of neurons in the middle temporal visual area (MT) of the primate cerebral cortex^{2,4}. In rhesus monkeys, these directionally selective 'pattern neurons' are distinguished by their ability to signal correctly the motion of a perceptually coherent plaid pattern^{2,4}. Directionally selective neurons in primary visual cortex (V1) and the majority of MT neurons ('component neurons'), by contrast, signal only the motions of the component gratings in a perceptually coherent plaid pattern^{2,4}. To investigate the contribution of area MT to visual motion coherence, we examined whether the responses of MT neurons can be modified by the same factors known to influence this perceptual decision. We used the phenomenon of perceptual transparency as a means to generate both coherent and noncoherent plaid patterns³ (Fig. 1). We studied 105 MT neurons in two rhesus monkeys. Directional tuning was first assessed using a drifting grating. The peak in this tuning curve defined the preferred direction for each cell. Directional selectivity was then examined using each of three different plaid configurations. Two of these stimuli were nontransparent and judged to move coherently by human observers³. The third plaid stimulus was transparent and judged to move noncoherently by human observers³.

Data obtained from a typical neuron are shown in Fig. 2a. When presented with either of the perceptually coherent plaids, this cell responded more strongly when the pattern moved in

the preferred direction (the 'pattern response') than when either of the components moved in this same direction (the 'component responses'). This type of directional selectivity (pattern response larger than component responses) is characteristic of pattern neurons^{2,4} and reflects integration of locally derived motion signals. When presented with the perceptually noncoherent plaid, however, this neuron's behaviour changed: the pattern response decreased by 42% and component responses were elevated slightly (by an average of 15%). These changes resulted in a type of direction tuning (component responses larger than pattern response) that is normally characteristic of component neurons in areas MT and V1 (refs. 2, 4). Hence the cell's responses became more component-like when the stimulus was configured to render noncoherent motion as the dominant percept.

Changes in directional selectivity that accompany perceptual changes were not restricted to neurons classified as pattern type. This can be illustrated effectively by subtracting responses obtained using coherent plaids from those obtained using noncoherent plaids. Difference curves of this sort are plotted in Fig. 2b for three additional MT neurons, which fell into component, pattern, and 'unclassifiable' categories using established criteria^{2,4}. Despite different classifications, all three neurons exhibited enhanced component responses and diminished pattern responses when stimulated with noncoherent plaids.

If these coherence-related changes in MT are to account for the perceptual phenomenon, then the responses of the neural population as a whole should become more component-like and less pattern-like in the presence of visual stimuli that are more apt to yield a percept of noncoherent motion. To test this prediction, we quantified coherence-related changes in directional tuning using an extension of a procedure introduced by Movshon *et al.*² (see legend to Figure 3). Briefly, predictions of both component- and pattern-type direction tuning to plaids were formed for each cell on the basis of responses to drifting gratings. The degree of similarity between observed plaid tuning curves and each of the component and pattern predictions was quantified by computing a measure of the variance accounted for by each prediction. Finally, to acquire a single index of how 'component-like' a neuron's tuning was for each stimulus condition, we computed the difference between the variance accounted for by the two predictions: the larger the value of this 'component index', the more component-like the response. The range and extent of coherence-related changes across our sample can be inspected by plotting the component index computed for each neuron under noncoherent conditions as a function of that obtained under each of the coherent conditions (Fig. 3a, b). Points falling above the diagonals in these plots reflect a shift toward component-like behaviour under stimulus conditions that usually yield a percept of component motion. Although there were varying degrees of coherence-related shifts, pairwise comparisons of the distributions revealed the component index to be systematically larger for the noncoherent condition than for either of the two coherent conditions (one-tailed paired *t*-test, $P < 0.001$).

The selective integration of locally derived motion signals is a perceptual phenomenon of profound behavioural significance. Retinal images with locally identical motions can render very different motion percepts, depending upon how the visual scene is parsed. Consider, for example, the dynamic retinal image elicited by the movements of a primate through a forest canopy. Moving image contours arise from a variety of real object boundaries as well as a combination of variations in surface reflectance, shadows, and occlusion. Accurate assessment of the motion of any of the objects in this scene (the primate, the swaying foliage, or perhaps the shadows cast by either) depends not only on integrating component motion signals but also on integrating only those signals common to each moving object. Our results suggest that the visual motion system accomplishes

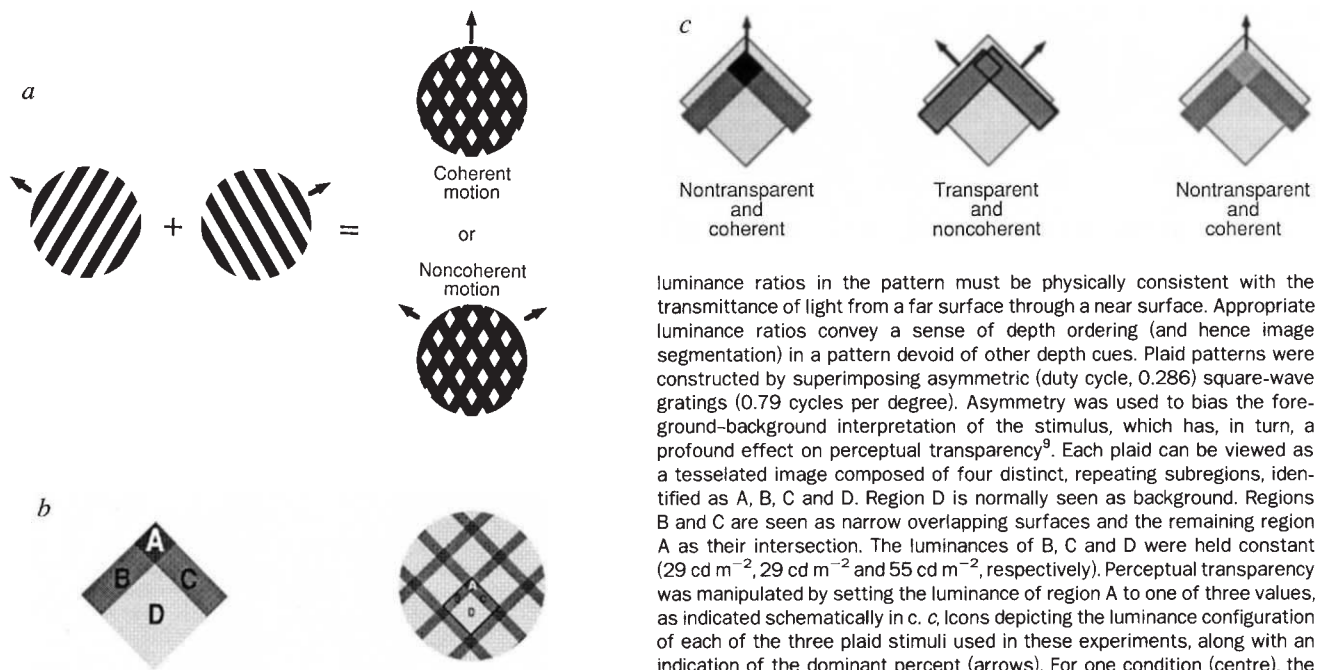


FIG. 1 *a*, Moving plaid patterns are produced by superposition of two drifting periodic gratings. The resultant pattern can be perceived to move either coherently or noncoherently depending on a variety of stimulus parameters^{1-3,5-9}. *b*, The procedures for creating perceptually transparent plaid patterns are derived from the physics of transparency^{10,11}. Simply put,

luminance ratios in the pattern must be physically consistent with the transmittance of light from a far surface through a near surface. Appropriate luminance ratios convey a sense of depth ordering (and hence image segmentation) in a pattern devoid of other depth cues. Plaid patterns were constructed by superimposing asymmetric (duty cycle, 0.286) square-wave gratings (0.79 cycles per degree). Asymmetry was used to bias the foreground-background interpretation of the stimulus, which has, in turn, a profound effect on perceptual transparency⁹. Each plaid can be viewed as a tessellated image composed of four distinct, repeating subregions, identified as A, B, C and D. Region D is normally seen as background. Regions B and C are seen as narrow overlapping surfaces and the remaining region A as their intersection. The luminances of B, C and D were held constant (29 cd m^{-2} , 29 cd m^{-2} and 55 cd m^{-2} , respectively). Perceptual transparency was manipulated by setting the luminance of region A to one of three values, as indicated schematically in *c*. Icons depicting the luminance configuration of each of the three plaid stimuli used in these experiments, along with an indication of the dominant perceived motion (arrows). For one condition (centre), the luminance of region A was chosen to be consistent with transparency (19 cd m^{-2}), yielding a percept of noncoherent motion in human observers³. For the remaining two conditions, the luminance of region A was either too dark (left, 3 cd m^{-2}) or too bright (right, 34 cd m^{-2}) to be compatible with transparency. These nontransparent plaids generally elicit a percept of coherent motion in human observers³. Other procedures for stimulus generation were as before³.

FIG. 2 Neural correlates of perceptual motion signal integration. *a*, Differential responses of an MT pattern-type neuron to coherent against noncoherent plaid patterns. Left, Directional tuning for a single drifting grating. Right, Responses to coherent and noncoherent plaids. Response amplitude (Δ spikes s^{-1}) was computed as the mean spike rate in the period of stimulus presentation minus the average baseline spike rate from all prestimulus periods. Error bars indicate s.e.m. response. When stimulated with either of the two types of perceptually coherent plaid patterns (filled circles, intersection region A too dark for transparency; open squares, intersection region A too bright for transparency; Fig. 1), response was maximal when the pattern moved in the neuron's preferred direction (0° in graph at right). But when stimulated with perceptually noncoherent plaid patterns (triangles), responses were maximal when either component moved in the preferred direction ($\pm 67.5^\circ$ in graph at right). *b*, Data from three additional MT neurons showing 'difference curves' (bottom row), computed by subtracting tuning curves obtained with coherent and noncoherent plaids (open squares and filled triangles, respectively, in top row). Pattern direction (P) is defined as the peak in the grating curve (filled squares, top row) and component directions (C) are displaced (67.5°) to each side. All three conventional neuronal types (component, pattern and unclassified) exhibited enhanced component responses and attenuated pattern responses when stimulated with noncoherent against coherent plaids. Visual stimuli were presented in a 10.5° circular aperture centred on the receptive field of each neuron. Gratings presented singly were moved at 10.26 degrees per s. Gratings composing plaid patterns were moved at 3.93 degrees per s and in directions differing by 135° , resulting in a pattern speed of 10.26 degrees per s. Stimuli were presented for five trials of 1.25 s exposure in each direction on a random schedule. Animals were alert and fixating during data acquisition. Eye position was monitored using a scleral search coil. Data from trials in which fixation deviated more than 0.5° were rejected.

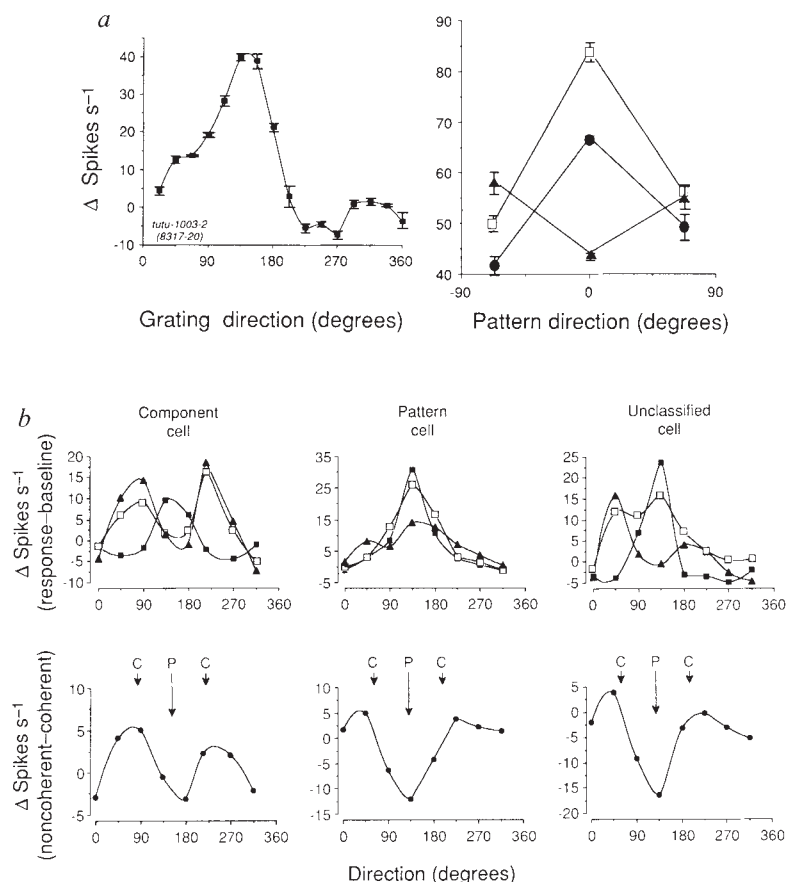


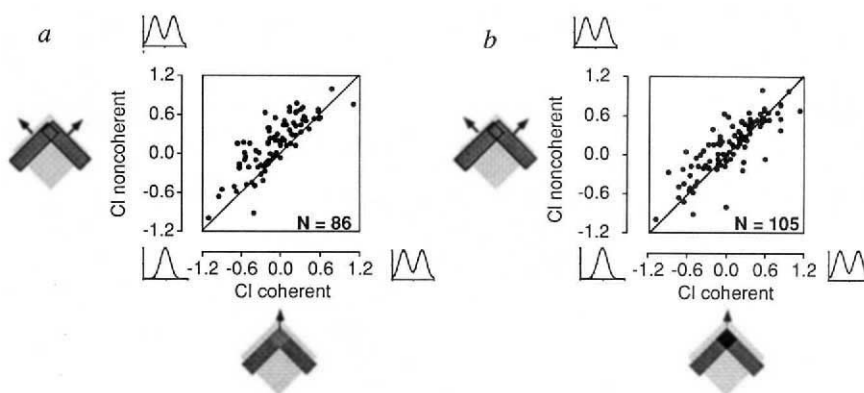
FIG. 3 Quantitative comparison of MT direction tuning for coherent against noncoherent plaid patterns. A 'component index' (CI) was computed by subtracting the variance accounted for by the pattern prediction from that accounted for by the component prediction. Each point in these graphs represents the value of the CI obtained for noncoherent (vertical axes) against one of the two coherent (horizontal axes) conditions. The larger the CI the more component-like (more bilobed, generally) the tuning curve. The graph at left compares CIs for the single noncoherent (transparent) case and one of the two coherent cases (nontransparent, region $A=34 \text{ cd m}^{-2}$; see Fig. 1). The graph at right compares CIs for the single noncoherent case and the other coherent case (region $A=3 \text{ cd m}^{-2}$; see Fig. 1). The tendency for points in both graphs to fall above the diagonal reflects a highly significant shift among our sample of MT neurons toward component-like responses in the presence of stimulus conditions that normally elicit a percept of noncoherent motion.

METHODS. The component index was computed using an extension of a procedure introduced by Movshon *et al.*² (see also ref. 4). Briefly, both a component and a pattern prediction were derived for each neuron from the observed responses to conventional drifting gratings. The actual direction tuning curves obtained using each plaid pattern were then compared with predicted component and pattern curves by computing the appropriate partial correlations as follows:

$$R_p = (r_p - r_{pc}r_c) / [(1 - r_c^2)(1 - r_{pc}^2)]^{0.5}$$

this selective integration by using segmentation cues, such as transparency, to classify motion signals according to physical origin. Furthermore, this selective integration is expressed in the behaviour of those neurons presumed to underlie perceptual integration of motion signals. Thus, an MT neuron that would have been classified as pattern-type using perceptually coherent plaids and established criteria^{2,4} will express component-type behaviour when stimulated with patterns that are more likely to result in a percept of noncoherent motion.

These results have two important implications. First, there is the implied correlation between neural and perceptual state, providing a first peek at the neural events underlying the selective integration of component motion signals. Although we have emphasized coherence-related shifts in the sample distribution, it remains to be seen whether behavioural reports of motion coherence are correlated on a trial-by-trial basis with the neuronal response from a single cell. Second, our data, together with mounting psychophysical evidence bearing on the factors that influence motion signal integration^{1-3,5-9}, demonstrate that figural aspects of a visual image unrelated to motion *per se* (such as perceptual transparency) have profound modulatory effects on the processing of motion signals in the primate visual system. We suggest that strict notions of modularity in visual processing must be modified to take these factors into account. □



where R_p = partial correlation for the pattern prediction, r_c = raw correlation of the data with the component prediction, r_{pc} = correlation of the two predictions. The partial correlation coefficient for the component prediction (R_c) was produced by exchanging r_p and r_c . The variance in each plaid direction tuning curve that can be accounted for by component and pattern predictions is thus R_c^2 and R_p^2 , respectively. The component index was computed as the difference between component and pattern variances ($R_c^2 - R_p^2$). CI is therefore a composite measure reflecting the degree to which a cell expresses component-type selectivity more strongly than pattern-type selectivity, for a specified stimulus condition.

Phosphorylation of transcription factor p62^{TCF} by MAP kinase stimulates ternary complex formation at c-fos promoter

Hendrik Gille, Andrew D. Sharrocks & Peter E. Shaw

Max-Planck-Institut für Immunbiologie, Spemann Laboratories, Postfach 1169, Stübeweg 51, W-7800 Freiburg, Germany

TRANSCRIPTION of the proto-oncogene c-fos is stimulated rapidly and transiently by serum growth factors and mitogens¹. Critical for this response is the serum-response element which is bound *in vivo* in a ternary complex containing the transcription factors p67^{SRF} and p62^{TCF} (ref. 2). Disruption of the ternary complex correlates with impaired induction by serum and phorbol ester^{3,4}. Mitogen-activated protein (MAP) kinase is a serine/threonine kinase which is activated 1–5 minutes after treatment of cells with mitogens and growth factors^{5–8} that induce re-entry into the cell cycle, making MAP kinase a candidate for the transmission of proliferative signals. Here we show that p62^{TCF} is phosphorylated by MAP kinase *in vitro* and that phosphorylation results in enhanced ternary complex formation. Serum-starved Swiss 3T3 cells treated with epidermal growth factor, which induces MAP kinase in these cells⁹, are induced to express c-fos and yield p62^{TCF} active in ternary complex formation. In contrast, treatment of Swiss 3T3 cells with insulin, which does not activate MAP kinase under these conditions⁹, does not lead to enhanced ternary complex formation nor does it induce c-fos transcription. Our results link the expression of the human c-fos proto-oncogene to signal transduction pathways known to be activated before its own induction.

We observed that ternary complex formation over the serum-response element (SRE) *in vitro* was more efficient with a truncated p67^{SRF} derivative (core^{SRF}, corresponding to residues 132–222 of p67^{SRF})¹⁰ from programmed rabbit reticulocyte lysates than it was with bacterially expressed core^{SRF}. As

Received 11 February; accepted 15 June 1992

- Adelson, E. A. & Movshon, J. A. *Nature* **300**, 523–525 (1982).
- Movshon, J. A., Adelson, E. H., Gizzi, M. & Newsome, W. T. in *Study Group on Pattern Recognition Mechanisms* (eds Chagas, C., Gattass, R. & Gross, C. G.) (Pontificia Academia Scientiarum, Vatican City, 1985).
- Stoner, G. R., Albright, T. D. & Ramachandran, V. S. *Nature* **344**, 153–155 (1990).
- Rodman, H. R. & Albright, T. D. *Exp Brain Res* **75**, 53–64 (1989).
- Krauskopf, J. & Farell, B. *Nature* **348**, 328–331 (1990).
- Kooi, F. L., DeValois, K. K. & Switkes, E. *Perception* (in the press).
- Vallortigara, G. & Bressan, P. *Vision Res.* **31**, 1967–1978 (1992).
- Adelson, E. A. *J. opt. Soc. Am. A* **1**, 1266 (1984).
- Stoner, G. R. & Albright, T. D. *Invest. Ophthalmol. Vis. Sci.* **33**, 1050 (1992).
- Metelli, F. *Scient. Am.* **230**, 91–95 (1974).
- Beck, J. in *Human and Machine Vision II* (ed. Rosenfeld, A.) 1–12 (Academic, New York, 1986).

ACKNOWLEDGEMENTS. We thank J. Costanza for technical assistance and V. S. Ramachandran for discussion. This work was supported by a grant from the NIH, a McKnight Foundation Development Award to T.D.A., and a Research Fellowship to G.R.S. from the McDonnell-Pew Center for Cognitive Neuroscience at San Diego.

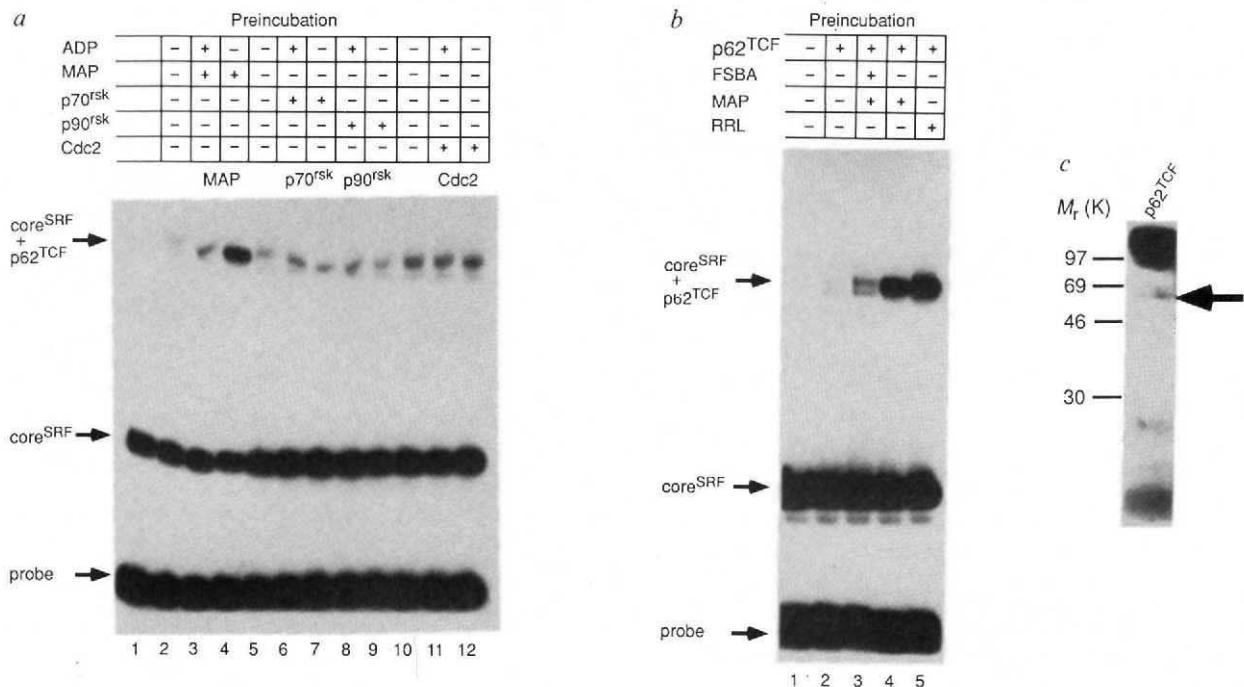


FIG. 1 p62^{TCF} phosphorylation and activation by MAP kinase. **a**, Activation of p62^{TCF} and ternary complex formation was assessed by gel retardation with bacterially expressed core^{SRF} and a ³²P-labelled SRE oligonucleotide probe (lane 1)³. Aliquots of an enriched fraction of p62^{TCF} were diluted in the different kinase buffers and reacted as indicated in the absence of kinase (lanes 2, 5, 10), with kinase and an excess of inhibitor (lanes 3, 6, 8, 11) or with kinase alone (4, 7, 9, 12) before addition of inhibitor. A slight increase in basal ternary complex formation was observed with the cdc2 kinase buffer (lanes 10–12). **b**, MAP kinase and reticulocyte lysate both activate p62^{TCF}. Activation and ternary complex formation was assayed with core^{SRF} as in **a**. p62^{TCF} was preincubated in MAP kinase buffer containing okadaic acid and 100 μ M ATP (lane 2) with MAP kinase, but in the presence of 5 mM 5'-*p*-fluorosulphonylbenzoyl adenosine (FSBA) (lane 3), with MAP kinase alone (lane 4) or with rabbit reticulocyte lysate (RRL) (lane 5). Inclusion of 5 mM FSBA results in 80% inhibition of the activation by reticulocyte lysate (not shown). **c**, A 62K protein (arrowed) present in the p62^{TCF} preparation is phosphorylated by MAP kinase, but in the presence of 100 μ M ATP, 50 μ M [γ -³²P]ATP and 0.2 mg ml⁻¹ histone H1. Incorporation of label was controlled by running an aliquot of the reaction mixture on an SDS gel. p34^{cdc2} activity was measured by adsorption to phosphocellulose paper³⁰. All kinases were active with their model substrates. DNA-binding reactions were set up as described with an SRE oligonucleotide³ and bacterially expressed core^{SRF} (A.D.S., unpublished results). p62^{TCF}, prepared as described¹⁰, was pretreated as indicated and added directly to the binding reactions. After incubation at room temperature for 10 min, samples were loaded onto 5% polyacrylamide gels. p62^{TCF} was labelled with MAP kinase in the presence of 100 μ M ATP and 10 μ M [γ -³²P]ATP at 30 °C for 30 min and analysed by SDS-PAGE.

METHODS. MAP kinase reactions were done in 25 mM Tris-HCl, pH 7.2, 10 mM MgCl₂, 1 mM dithiothreitol (DTT), 0.1 mM EGTA, 0.1 mM sodium

orthovanadate, 1 μ M okadaic acid, 100 μ M ATP, with 5–10 units of enzyme (purified as described⁹) for 20 min at 37 °C. Kinase reactions were terminated by addition of 5'-*p*-fluorosulphonylbenzoyl adenosine to 5 mM or ADP to 25 mM. MAP kinase activity was assayed with myelin basic protein (0.33 mg ml⁻¹) as described²⁹. p70^{S6k} and p90^{rsk} reactions were carried out in 50 mM MOPS, pH 7.2, 1 mM DTT, 10 mM MgCl₂, 0.05% Triton X-100, 100 μ M ATP, 10 mM *p*-nitrophenylphosphate. p90^{rsk} activity was measured in buffer containing 100 μ M ATP, 20 μ M [γ -³²P]ATP and 100 μ M Kemptide as the phosphate acceptor. Incorporation was measured by spotting 10- μ l aliquots of the reaction onto P81 phosphocellulose paper³⁰. Cdc2 kinase reactions were done in 50 mM HEPES/KOH, pH 7.4, 10 mM MgCl₂, 1 mM DTT and 0.5 mM ATP. Cdc2 activity was measured in the presence of 100 μ M ATP, 50 μ M [γ -³²P]ATP and 0.2 mg ml⁻¹ histone H1. Incorporation of label was controlled by running an aliquot of the reaction mixture on an SDS gel. p34^{cdc2} activity was measured by adsorption to phosphocellulose paper³⁰. All kinases were active with their model substrates. DNA-binding reactions were set up as described with an SRE oligonucleotide³ and bacterially expressed core^{SRF} (A.D.S., unpublished results). p62^{TCF}, prepared as described¹⁰, was pretreated as indicated and added directly to the binding reactions. After incubation at room temperature for 10 min, samples were loaded onto 5% polyacrylamide gels. p62^{TCF} was labelled with MAP kinase in the presence of 100 μ M ATP and 10 μ M [γ -³²P]ATP at 30 °C for 30 min and analysed by SDS-PAGE.

reticulocyte lysates contain active protein kinases, a partially purified fraction of p62^{TCF} (ref. 10) was incubated with various protein kinases and analysed for ternary complex formation by gel retardation. We tested kinases that are rapidly activated following mitogenic stimulation of quiescent cells, including p70^{S6k} (refs 11, 12), p90^{rsk} (ref. 13) and MAP kinases^{5–9}. In addition we investigated whether p34^{cdc2} kinase, which is also activated at the G0–G1 transition¹⁴, would phosphorylate p62^{TCF}. Of these, only MAP kinase treatment of p62^{TCF} resulted in enhanced ternary complex formation with core^{SRF} *in vitro* (Fig. 1a). A concomitant change in ternary complex mobility was observed, arguing for a direct modification of p62^{TCF}. Stimulation of ternary complex formation by MAP kinase could be inhibited by the addition of 5'-*p*-fluorosulphonylbenzoyl adenosine, a reagent used to modify the nucleoside-binding site of kinases and render them inactive¹⁵ (Fig. 1b). Addition of the competitive inhibitor ADP also blocked the reaction (Fig. 1a). Ternary complexes were not formed with SRE oligonucleotides containing mutations that specifically prevent p62^{TCF} binding but leave p67^{SRF} binding unaffected³ (data not shown). Consistent with our initial inference, pre-treatment of p62^{TCF} with

either reticulocyte lysate or purified MAP kinase has the same qualitative and quantitative effect on ternary complex formation (Fig. 1b). To test whether p62^{TCF} is modified directly, we incubated a partially purified fraction with MAP kinase in the presence of [γ -³²P]ATP and analysed the reaction products by SDS-polyacrylamide gel electrophoresis. A protein with an apparent relative molecular mass (*M_r*) of 62,000 (62K) was labelled (Fig. 1c).

An activity augmented by phosphorylation should also be susceptible to phosphatase treatment. Incubation of p62^{TCF} with potato acid phosphatase in the presence of phosphatase inhibitors did not affect p62^{TCF} binding, but addition of the acid phosphatase before the inhibitors abolished ternary complex formation (Fig. 2a, lane 3). Subsequent incubation of acid phosphatase-treated p62^{TCF} with MAP kinase in the presence of phosphatase inhibitors restored ternary complex formation, confirming the integrity of acid-phosphatase-treated p62^{TCF} (Fig. 2, lane 4). Treatment of MAP kinase-activated p62^{TCF} with acid phosphatase abolished complex formation (Fig. 2, lane 6), emphasizing the reversibility of the effect.

Full-length p67^{SRF} purified from HeLa cells was also tested

as a potential substrate for the various kinases. Notably, MAP kinase did not phosphorylate p67^{SRF} and had no effect on p67^{SRF} binding under the same conditions used to phosphorylate and test p62^{TCF} (data not shown). Only incubation with p90^{rsk} led to a slight enhancement of p67^{SRF} binding to the SRE, but this did not lead to increased ternary complex formation (data not shown).

MAP kinase is rapidly activated in Swiss 3T3 cells upon treatment with epidermal growth factor (EGF), whereas, in the same cells, insulin stimulates the activity of p70^{S6k} but not MAP kinase⁹. We therefore assessed the effect of these two agonists on the expression of *c-fos* in serum-starved Swiss 3T3 cells. Northern analysis of RNA isolated from starved cells before and after treatment with either EGF or insulin revealed that insulin fails to stimulate *c-fos* transcription, whereas EGF induces the gene rapidly (Fig. 3).

If phosphorylation of p62^{TCF} by MAP kinase and the consequent enhancement of ternary complex formation are part of the signal cascade leading to rapid induction of *c-fos* expression, enhanced ternary complex formation ought to be detectable in Swiss 3T3 cells treated with EGF but not in insulin-treated or untreated cells. To address this question, nuclear extracts were prepared from serum-starved Swiss 3T3 cells and from serum-starved cells subsequently treated with EGF or insulin for various times before harvesting. Each extract was tested for its ability to produce the ternary complex with the SRE. Treatment of starved cells with EGF for 15 and 30 min increased the yield

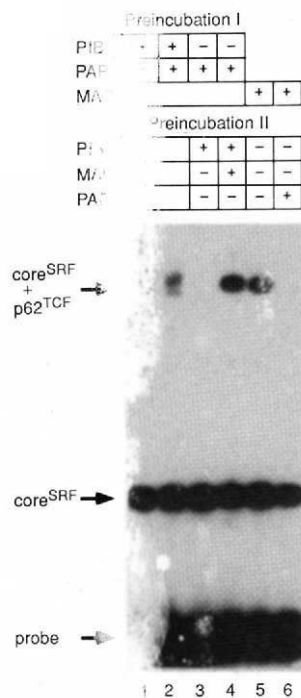


FIG. 2 Ternary complex formation by p62^{TCF} is susceptible to phosphatase. Aliquots of a partially purified preparation of p62^{TCF} were pretreated sequentially, first with either potato phosphatase (PAP) or MAP kinase (preincubation I) and then with MAP kinase or PAP (preincubation II), as indicated. Reactions were subsequently assayed to preformed complexes of core^{SRF} and an SRE oligonucleotide probe. p62^{TCF} was incubated in MAP kinase buffer in the presence of phosphatase inhibitors (PIB) (lane 1) or with phosphatase inhibitors and 1.8 × 10⁻² units of PAP (lane 2), or with phosphatase alone (lane 3). After phosphatase treatment and addition of PIB, p62^{TCF} was incubated with MAP kinase (lane 4). p62^{TCF} activated by MAP kinase (lane 5) was also treated with 0.15 units of PAP (lane 6). METHODS. PAP was prepared as described¹³. Phosphatase reactions were carried out in the same buffer as the MAP kinase reactions because PAP activity was largely insensitive to buffer conditions. Phosphatase inhibitor buffer consists of 1 mM NaVO₄, 10 mM *p*-nitro-phenylphosphate, and 10 mM β -glycerophosphate¹³.

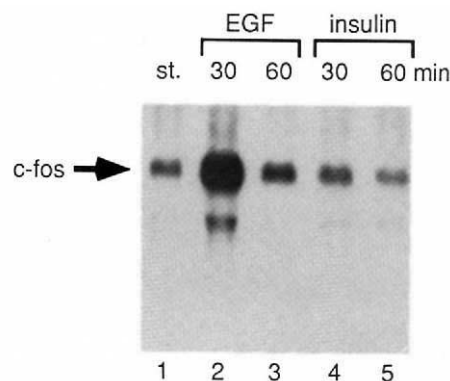


FIG. 3 Treatment of Swiss 3T3 cells with EGF but not with insulin induces *c-fos* expression. RNA was isolated from starved cells (st.; lane 1) or from cells treated for 30 and 60 min with EGF (lanes 2 and 3), or with insulin (lanes 4 and 5) and analysed by northern blotting. METHODS. Swiss 3T3 cells were starved and stimulated with 5 nM EGF or 1 nM insulin as described³¹. Total RNA was isolated after cell lysis in guanidinium thiocyanate and 8 μ g of each sample was electrophoresed in a 1% agarose/formaldehyde gel. RNA was transferred to a nylon membrane and hybridized with a radiolabelled *c-fos* riboprobe.

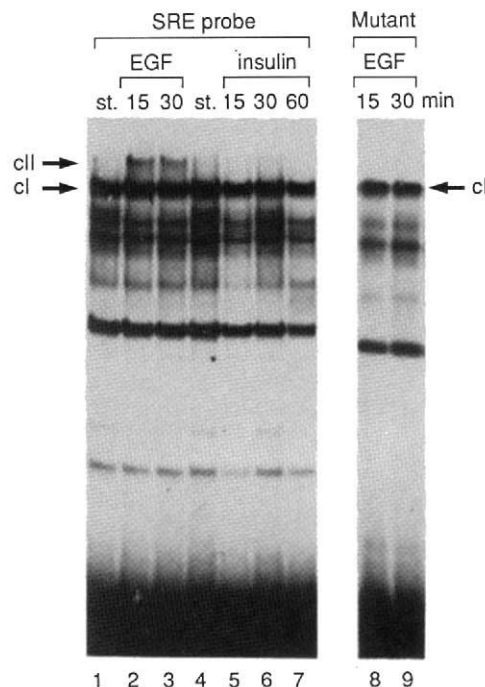


FIG. 4 Increased ternary complex formation is detected after stimulation of Swiss 3T3 cells with EGF but not after insulin treatment. Ternary complex formation was assessed by gel retardation: cI refers to the complex between p67^{SRF} and the SRE, cII refers to the ternary complex. Nuclear extracts were prepared from starved cells (st; lane 1 and 4) or from cells stimulated for 15 or 30 min with EGF (lanes 3 and 4) or insulin for 15, 30 and 60 min (lanes 5–7). The ternary complex seen after treatment with EGF for 15 or 30 min is not formed with the probe 'el' (ref. 3), which carries mutations that specifically inhibit ternary complex formation by p62^{TCF} (lanes 8 and 9). METHODS. Swiss 3T3 cells were starved and stimulated as for RNA preparation. Cells were collected from 150-mm dishes on ice and nuclear extracts were prepared by a modification of the procedure described in ref. 3. Briefly, cells were lysed in buffer A by passage through a hypodermic needle. Nuclei were collected by low-speed centrifugation and eluted in buffer C by stirring on ice for 45 min. The eluate was dialysed against buffer D in a microdialysis unit. Buffers were prepared as described³ except for the addition of 1 mM NaVO₄, 5 mM NaF, 20 mM β -glycerophosphate and 10 mM pNPP to buffers A and C. In addition, buffer C was supplemented with 200 nM okadaic acid. The reproducibility of the effect described depends on the presence of these phosphatase inhibitors in the buffers used for the preparation of extracts.

of ternary complex, but treatment with insulin for up to 60 min had no effect (Fig. 4). The ternary complex was susceptible to mutations in the probe that specifically prevent p62^{TCF} binding but leave p67^{SRF} binding unaffected³. Enhanced formation of a similar complex over the *c-fos* SRE has also been observed in a human astrocytoma cell line upon treatment with EGF and other mitogens¹⁶. We infer that these effects are caused by phosphorylation of p62^{TCF} by MAP kinase. *In vivo* footprint analyses have indicated that occupation of the SRE by transcription factors is unchanged upon growth factor stimulation of *c-fos* expression, although a rapid exchange of factors could not be ruled out². This latter interpretation is more in accord with the implications of our data.

Phosphorylation of p67^{SRF} by casein kinase II is considered to be an important event in the regulation of *c-fos* expression¹⁷⁻¹⁹. But it is unlikely to be involved in induction, as casein kinase II is activated by growth factors after 15-30 min, a period in which *c-fos* expression has already reached its peak^{20,21}. Furthermore, the p67^{SRF} tryptic and chymotryptic phosphopeptide patterns do not change upon stimulation of cells (ref. 22; and H. König, personal communication). In contrast, active nuclear MAP kinase can be detected as early as 5 min after stimulation of cells with growth factors²³.

Two independently isolated cDNA clones have been reported to encode proteins that have the p67^{SRF}-dependent DNA-binding characteristic of p62^{TCF} (refs 24, 25). Both proteins, Elk-1 and Sap-1 contain the amino-acid motifs PL(S/T)P and PR(S/T)P in their C termini. These motifs have been shown to be phosphorylated efficiently by MAP kinase using *in vivo* substrates or synthetic peptides^{26,27}. This is consistent with the transcription factor p62^{TCF} being encoded by either *elk-1* or *sap-1*. Indeed, a C-terminal fragment of Elk-1 can be phosphorylated by MAP kinase *in vitro* (H.G., unpublished).

Fos induction by serum and phorbol ester requires p62^{TCF} binding. As stimulation of *c-fos* transcription does not require protein synthesis, it is likely to be brought about by the modification of existing factors²⁸. From our results we conclude that MAP kinase, which can be rapidly activated in resting cells, is located directly upstream from p62^{TCF} in the intracellular signalling pathway that leads to *c-fos* induction. □

Raf-1 activates MAP kinase-kinase

John M. Kyriakis, Harald App*, Xian-feng Zhang, Papia Banerjee, David L. Brautigan†, Ulf R. Rapp* & Joseph Avruch

Diabetes Unit, Medical Services, Massachusetts General Hospital, 149 13th Street, Charlestown, Massachusetts 02129, and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA

*Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA

†Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912, USA

THE normal cellular homologue of the acutely transforming oncogene *v-raf* is *c-raf-1*, which encodes a serine/threonine protein kinase that is activated by many extracellular stimuli¹. The physiological substrates of the protein *c-Raf-1* are unknown. The mitogen-activated protein (MAP) kinases Erk1 and 2 are also activated by mitogens through phosphorylation of Erk tyrosine and threonine residues catalysed by a protein kinase of relative molecular mass 50,000, MAP kinase-kinase (MAPK-K)²⁻⁷. Here we report that MAPK-K as well as Erk1 and 2 are constitutively active in *v-raf*-transformed cells. MAPK-K partially purified from *v-raf*-transformed cells or from mitogen-treated cells³ can be deactivated by phosphatase 2A. *c-Raf-1* purified after mitogen stimulation can reactivate the phosphatase 2A-inactivated MAPK-K over 30-fold *in vitro*. *c-Raf-1* reactivation of MAPK-K coincides with the selective phosphorylation at serine/threonine residues of a polypeptide with *M_r* 50,000 which coelutes precisely on cation-exchange chromatography with the MAPK-K activatable by *c-Raf-1*. These results indicate that *c-Raf-1* is an immediate upstream activator of MAPK-K *in vivo*. To our knowledge, MAPK-K is the first physiological substrate of the *c-raf-1* protooncogene product to be identified.

We explored the possibility that Raf-1 kinases lie upstream from MAP kinases in the mitogen signalling cascade using two cell lines: F4, an NIH3T3 cell stably transformed with *v-raf*, and CL-7, the parental NIH3T3 clone. Cytosolic extracts of serum-starved cells were fractionated over Mono-Q anion-exchange columns. Fractions were assayed for their ability to phosphorylate myelin basic protein (MBP), a common MAP kinase substrate⁵. Although CL-7 cell crude extracts and Mono-Q fractions contain levels of Erk1/2 immunoreactivity that are the same as those in F4 cell extracts (Fig. 1b, top panel, and c), serum-starved unstimulated CL-7 cells have little or no MBP kinase activity (Fig. 1a, top; open symbols). By contrast, serum-starved F4 cells, in the absence of agonist stimulation, display a vigorous endogenous MBP kinase which resolves as two peaks (Fig. 1a, top; filled symbols). These F4 cell MBP kinase peaks contain Erk1/2 MAP kinase immunoreactive bands corresponding to relative molecular masses of 44,000 and 42,000, which comigrate with activity (Fig. 1b, second panel).

We tested the same Mono-Q fractions shown in Fig. 1a for their ability to activate a bacterially expressed inactive form of p44 MAP kinase/Erk1 (MAPK-K assay). Extracts from CL-7 cells contain little detectable MAPK-K (Fig. 1a, bottom, open symbols). F4 cells, however, contain marked MAPK-K activity which resolves from the endogenous MAP kinases. This MAPK-K is not itself the *v-Raf* protein. Immunoblotting of the fractions shown in Fig. 1a with anti-Raf-1 antibodies reveals a band corresponding to *M_r* 48,000 (48K), present only in the F4 cells, which elutes late in the gradient (Fig. 1b; compare third and fourth panels). This represents an active proteolytically processed product of *gag v-Raf* which can appear in extracts of retrovirally transformed cells such as F4 overexpressing *v-raf*⁸.

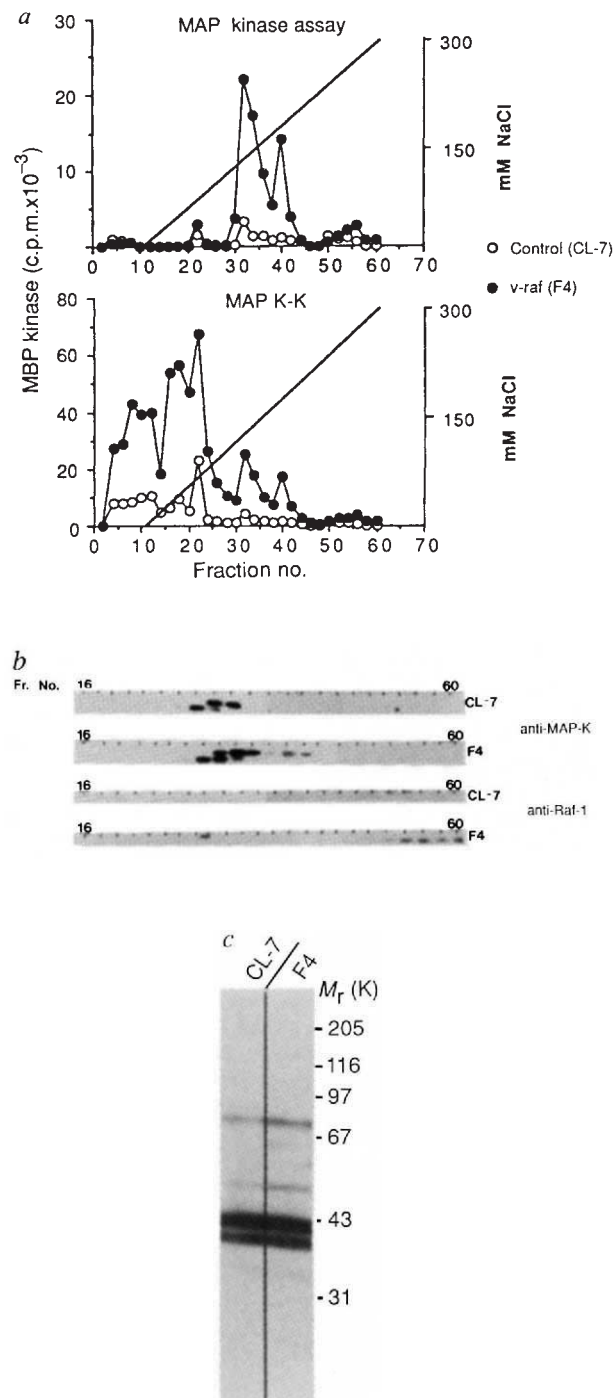
Received 23 March; accepted 8 June 1992.

- Herschman, H. R. A. *Rev. Biochem.* **60**, 281-319 (1991).
- Herrera, R. E., Shaw, P. E. & Nordheim, A. *Nature* **340**, 68-70 (1989).
- Shaw, P. E., Schröter, H. & Nordheim, A. *Cell* **56**, 563-572 (1989).
- Graham, R. & Gilman, M. *Science* **251**, 189-192 (1991).
- Ray, L. B. & Sturgill, T. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1502-1506 (1987).
- Hoshi, M., Nishida, E. & Sakai, H. *J. biol. Chem.* **263**, 5396-5401 (1988).
- Boulton, T. G. *et al. Cell* **65**, 663-675 (1991).
- Cobb, M. H., Boulton, T. G. & Robbins, D. J. *Cell Reg.* **2**, 965-978 (1991).
- Ballou, L., Luther, H. & Thomas, G. *Nature* **349**, 348-350 (1991).
- Schröter, H. *et al. EMBO J.* **9**, 1123-1130 (1990).
- Susa, M. & Thomas, G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7040-7044 (1990).
- Susa, M. & Thomas, G. *Cell* **57**, 817-824 (1989).
- Chen, R.-H. & Blenis, J. *Molec. cell. Biol.* **10**, 3204-3215 (1990).
- Hall, F. L., *et al. J. biol. Chem.* **266**, 17430-17440 (1991).
- Buhrow, S. A., Cohen, S., Garbers, D. L. & Staros, J. V. *J. biol. Chem.* **258**, 7824-7827 (1983).
- Malik, R. K., Roe, M. W. & Blackshear, P. J. *J. biol. Chem.* **266**, 8576-8582 (1991).
- Manak, J. R., deBisschop, N., Kris, R. M. & Prywes, R. *Genes Dev.* **4**, 955-967 (1990).
- Manak, J. R. & Prywes, R. *Molec. cell. Biol.* **11**, 3652-3659 (1991).
- Marais, R. M., Hsuan, J. J., McGuigan, C., Wynne, J. & Treisman, R. *EMBO J.* **11**, 97-105 (1992).
- Ackermann, P. & Osheroff, N. *J. biol. Chem.* **264**, 11958-11965 (1989).
- Ackermann, P., Glover, C. V. C. & Osheroff, N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 821-825 (1990).
- Prywes, R., Dutta, A., Cromlish, J. A. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7206-7210 (1988).
- Chen, R.-H., Sarnecki, C. & Blenis, J. *Molec. cell. Biol.* **12**, 915-927 (1992).
- Hipskind, R. A., Rao, V. N., Mueller, C. G. F., Reddy, E. S. P. & Nordheim, A. *Nature* **354**, 531-534 (1991).
- Dalton, S. & Treisman, R. *Cell* **68**, 597-612 (1992).
- Alvarez, E. *et al. J. biol. Chem.* **266**, 15277-15285 (1991).
- Gonzalez, F. A., Raden, D. L. & Davis, R. I. *J. biol. Chem.* **266**, 22159-22163 (1991).
- Greenberg, M. E., Green, L. A. & Ziff, E. *J. biol. Chem.* **260**, 14101-14110 (1985).
- Gomez, N., Tonks, N. K., Morrison, C., Harnar, T. & Cohen, P. *FEBS Lett.* **271**, 119-122 (1991).
- Ferrari, S. & Thomas, G. *Meth. Enzym.* **200**, 159-169 (1991).
- Ballou, L., Jenö, P. & Thomas, G. *J. biol. Chem.* **263**, 1188-1194 (1988).

ACKNOWLEDGEMENTS. We thank G. Thomas for S6 kinases and MAP kinase, M. Cobb for *erk2* (MAP) kinase and G. Draetta for p34^{cdc2} kinase; H. Schröter for a fraction of partially purified p62^{TCF}; P. Sassone-Corsi, S. Ferrari, P. Nielsen, M. Reth and G. Thomas for discussions and critical comments on the manuscript; and C. Loomis for an introduction to protein kinase biochemistry (H.G.). This work was supported by the Max-Planck-Gesellschaft and the Deutsche Forschungsgemeinschaft.

FIG. 1 MAP kinase and MAP kinase-kinase are activated in *v-raf*-transformed NIH 3T3 cells. *a*, Top, MAP kinase assay of basal F4 (*v-raf*-transformed) cell Mono-Q fractions (filled circles) or basal CL-7 (control) cells (open circles). Bottom, MAP kinase-kinase (MAPK-K) assay of the same samples. One-ml fractions were assayed. The salt gradient is indicated by the diagonal line and the right-hand ordinate. *b*, Immunoblots of Mono-Q even-numbered fractions (16–60) in *a*. From top, CL-7 cells, anti-MAP kinase; F4 cells, anti-MAP kinase antibody (asterisks indicate fractions with peak MBP-phosphorylating activity); CL-7 cells, anti-Raf-1; F4 cells, anti-Raf-1. *c*, Immunoblot of CL-7 and F4 cell crude extracts with anti-MAP kinase. M_r calibration shown on the right.

METHODS. Serum-starved CL-7 or F4 cells were rinsed twice with ice-cold PBS buffer. Thereafter all procedures were carried out at 4 °C. Cells (15 plates of CL-7 or 4 plates of F4) were scraped into 6 ml Mono-Q buffer (50 mM β -glycerophosphate, pH 7.3, 1.5 mM Na-EGTA, 100 μ M Na_2VO_4 , 1 mM dithiothreitol (DTT), 400 μ M PMSF, 2 μ M leupeptin, 2 μ M pepstatin, 10 kallikrein-inhibiting units per ml aprotinin). Cells were broken by Dounce homogenization (20 strokes), centrifuged (100,000g for 1 h) and the supernatants normalized to the same protein concentration (3.5 mg ml⁻¹). CL-7 and F4 cell extract supernatants (21 mg protein each) were then loaded onto Mono-Q HR 5/5 columns equilibrated with Mono-Q buffer. Columns were washed with 4 ml buffer and eluted with a 50-ml linear NaCl gradient (0–300 mM) in Mono-Q buffer. One-ml fractions of the flow-through, wash and eluate were collected. The flow rate was kept at 1 ml per min. Mono-Q column (10- μ l) fractions were mixed with 6 μ l 35 μ g ml⁻¹ bacterial recombinant p44^{erk1}/MAP kinase (bt MAP kinase) in buffer D (20 mM HEPES, pH 7.4, 1 mM Na-EGTA, 1 mM DTT, 0.4 mg ml⁻¹ crystallized/lyophilized BSA) (MAPK-K assays) or buffer D alone (MAP kinase assays) plus 4 μ l MgCl_2 /[γ -³²P]ATP (50 mM and 0.5 mM respectively; 2,400 c.p.m. pmol⁻¹). Samples were incubated for 30 min at 30 °C, at which time 10 μ l reaction mix (150 mM Tris-HCl, pH 7.4, 3 mM DTT, 30 mM MgCl_2 , 150 μ M [γ -³²P]ATP (2,400 c.p.m. pmol⁻¹), 0.75 mg ml⁻¹ MBP, 7.5 μ M heat-stable inhibitor of cAMP-dependent protein kinase²¹) were added. The incubation was continued for another 15 min, after which 15 μ l were spotted onto 2 cm $\times$ 2 cm squares of phosphocellulose paper. The papers were washed and counted by liquid scintillation spectroscopy. The btMAP kinase is a nearly full-length rat *erk1* expressed in *E. coli* as a glutathione S-transferase Erk1 fusion protein, purified by adsorption to glutathione-Sepharose followed by thrombin cleavage. Details of the cloning, expression and purification of btMAP kinase will be presented elsewhere. Immunoblotting with anti-Raf-1 antibody SP63 (raised against the C-terminal 12 amino acids of c-Raf-1) as well as anti-MAP kinase antibodies (raised against the bacterially expressed btMAP kinase) has been described²².



In addition, F4 cells do not contain more total (inactive plus active) MAPK-K than CL-7 cells, nor are CL-7 cells defective in mitogen signalling. We pretreated both cell lines with vanadate (200 μ M) and stimulated them with phorbol myristate acetate (1 μ M). Mono-Q fractions were obtained as for Fig. 1*a*. After mitogen stimulation both cells show comparable MAPK-K and MAP kinase activities (Fig. 2*a*).

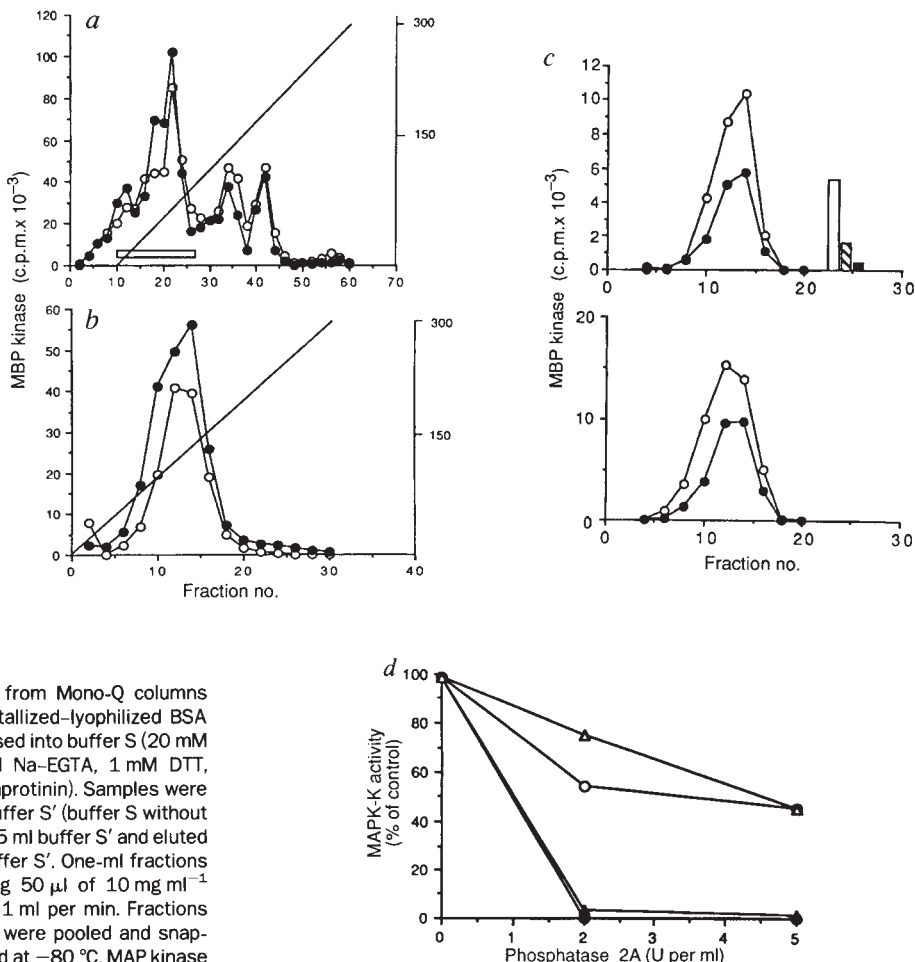
The chromatographic, catalytic and regulatory properties of the activated MAPK-K present in *v-raf*-transformed NIH3T3 cells resemble closely those of the MAPK-K from mitogen-stimulated CL-7 cells. MAPK-Ks from both cells coelute from Mono-Q and from Mono-S columns (Fig. 2*a* and *b*). Moreover, both the F4 cell and CL-7 cell MAPK-Ks activate not only the totally dephosphorylated bacterial MAP kinase but can reactivate purified H4 cell MAP kinases that have been deactivated by selective dephosphorylation at tyrosine or serine/threonine

residues (Fig. 2*c*). In both cases, Tyr and Ser/Thr-specific reactivating activities coelute from Mono-S at a constant activity ratio (Fig. 2*c*). Finally, both the F4 and the CL-7 cell MAPK-Ks can be completely and specifically inactivated with phosphatase 2A (Fig. 2*d*), confirming that both MAPK-Ks are regulated by Ser/Thr phosphorylation, and that the active MAPK-K in *v-raf*-transformed cells is active by virtue of Ser/Thr phosphorylation of the MAPK-K polypeptide.

The constitutive activation of the F4 cell MAPK-K impelled us to test whether Raf-1 itself was capable of phosphorylating MAPK-K directly *in vitro* and (re)activating its catalytic function. MAPK-K purified from insulin and vanadate-treated H4 hepatoma cells was devoid of endogenous MBP kinase after the Mono-S step and activated bacterial MAP kinase over 20-fold (Fig. 3*a*). Although not completely pure, the specific activity of the most highly purified fractions from Mono-S

FIG. 2 Comparison of MAPK-K from phorbol myristate acetate(PMA)/vanadate-treated, control or *v-raf*-transformed NIH3T3 cells. CL-7 (parental control) and F4 (*v-raf*-transformed) cells were treated with Na_3VO_4 (0.2 mM, 1 h) followed by PMA (1 μM , 20 min). *a*, Extracts were prepared, chromatographed on Mono-Q and 1-ml fractions assayed for MAPK-K activity as described. White bar indicates the peaks of MAPK-K activity; open symbols, CL-7 cells; filled symbols, F4 cells. *b*, Fractions containing MAPK-K activity (*a*) were pooled and chromatographed on Mono-S; open symbols, CL-7 cells; closed symbols, F4 cells. *c*, Mono-S fractions from *b* were tested for their ability to reactivate purified MAP kinases that had first been inactivated with either recombinant tyrosine phosphatase (PTP-1, open symbols), or (Ser/Thr) phosphatase 2A (filled symbols). Top, MAPK-K fractions from CL-7 cells; bottom, MAPK-K fractions from F4 cells. Bar graphs denote starting activity (open), activity after PTP-1 (hatched) and activity after phosphatase 2A (filled) measured according to the ordinate. *d*, Phosphatase 2A inactivation of MAPK-Ks from F4 cells (triangles) or CL-7 cells (circles). Okadaic acid (1 μM) was used during phosphatase incubations as indicated (open symbols).

METHODS. For Mono-S chromatography, fractions from Mono-Q columns containing MAPK-K activity were pooled, and crystallized-lyophilized BSA was added to 0.4 mg ml^{-1} . Samples were then dialysed into buffer S (20 mM MES, pH 6.5, 12.5 mM β -glycerophosphate, 1 mM Na-EGTA, 1 mM DTT, 400 μM PMSF, 10 kallikrein-inhibiting units per ml aprotinin). Samples were then loaded onto a Mono-S column equilibrated in buffer S' (buffer S without β -glycerophosphate). The column was washed with 5 ml buffer S' and eluted with a 30-ml linear NaCl gradient (0–300 mM) in buffer S'. One-ml fractions of the eluate were collected into tubes containing 50 μl of 10 mg ml^{-1} crystallized-lyophilized BSA; flow rate was kept at 1 ml per min. Fractions with MAPK-K activity (eluting at 90–170 mM NaCl) were pooled and snap-frozen in liquid nitrogen as 0.1-ml aliquots and stored at -80°C . MAP kinase was purified from insulin/vanadate-treated H4 hepatoma cells²³. Inactivation of purified H4 cell MAP kinases with purified phosphatase 2A or recombinant rat brain protein tyrosine phosphatase 1 was as described^{23,25}. In each case we used 0.05 U (1 U transfers 1 pmol min^{-1} PO_4 to MBP) MAP kinase activity, 250 U per ml PTP-1, and 5 U per ml phosphatase 2A (see refs 24 and 25



was 3,333 U per mg (where 1 U MAPK-K-activated MAP kinase transfers 1 pmol of phosphate per min from ATP to MBP); by comparison, the specific activity of purified *Xenopus* MAPK-K was reported to be 7,000 U per mg using a similar assay system⁹. Treatment of each Mono-S fraction with phosphatase 2A reduced MAPK-K activity by over 95% (Fig. 3*b*, triangles).

A constitutively active c-Raf-1, truncated to remove the N-terminal regulatory region was purified by immunoprecipitation from NIH BXB cells, which overexpress this mutant-active c-Raf-1. The c-Raf-1 immune precipitate is devoid of endogenous MBP kinase (Fig. 3*c*, *d*, *e*, bar F) and MAPK-K (Fig. 3*c*, *d*, *e*, bar E). Incubation of the Mono-S fractions containing phosphatase 2A-treated MAPK-K with the immunoprecipitated c-Raf results in a marked increase in MAPK-K activity, exceeding 30-fold in each fraction (Fig. 3*b*; compare open triangles to closed circles). The level of MAPK-K activity achieved by incubation of phosphatase-treated MAPK-K with c-Raf-1 and Mg-ATP actually exceeds the MAPK-K activity before phosphatase 2A treatment (Fig. 3*b*; compare filled circles with open circles). In fact, c-Raf-1 can further activate MAPK-K which has not been first subjected to phosphatase treatment (Fig. 3*c*; compare bars D and A). We attribute this 'superactivation' to the copurification of latent, inactive MAPK-K with active enzyme. In support of this conclusion, the c-Raf-1 immunoprecipitate is devoid of active or latent MAPK-K-activatable MBP kinase (Fig. 3*c*, bars F and H) or MAPK-K (bar E), and the MAPK-K isolates lack active (bar G) or latent

for unit definitions). MAPK-K purified through the Mono-S step was treated with purified phosphatase 2A at the concentrations indicated for 45 min at 30°C . Phosphatase inactivation was stopped with okadaic acid (1 μM) and MAPK-K activity measured.

Raf-activatable (bar H) MBP kinase. Thus the latent MAPK-K that is activated on incubation of MAPK-K isolates with c-Raf-1 (Fig. 3*c*; compare bars D and A) must be present initially in the MAPK-K isolate. Moreover, once this latent MAPK-K has been activated by c-Raf-1, it has regulatory properties indistinguishable from those of the MAPK-K purified in an active form (Fig. 3*d*). Like the active MAPK-K generated by mitogen treatment *in situ*, the c-Raf-1-activated MAPK-K (Fig. 3*d*, bar B) is also extensively deactivated by phosphatase 2A (Fig. 3*d*, bar C), and a second incubation with c-Raf-1 and Mg-ATP produces a second cycle of MAPK-K reactivation, but now ATP produces only the level before phosphatase treatment (Fig. 3*d*, bar D). The c-Raf-1-induced activation of MAPK-K requires ATP (Fig. 3*e*; compare bars B, C, D). Analysis of Mono-S fractions incubated in the presence of Mg^{2+} and $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ by SDS-polyacrylamide gel electrophoresis and autoradiography demonstrates that addition of c-Raf-1 leads to the selective phosphorylation of a 50K polypeptide (Fig. 3*f*). This c-Raf-1 substrate coelutes precisely with the peak of c-Raf-1-activated MAPK-K. Phosphoamino-acid analysis of the 50K polypeptide reveals that levels of ^{32}P -serine and ^{32}P -threonine are nearly the same (Fig. 3*g*). Based on the coincident elution positions of the 50K c-Raf-1 substrate and c-Raf-activated MAPK-K activity on cation-exchange chromatography, we propose that the 50K c-Raf-1 polypeptide substrate is likely to be the c-Raf-1-activated MAPK-K. This conclusion is consistent with the identification of *Xenopus* MAPK-K as a 50K polypeptide⁹. These data indicate that c-Raf-1 directly phosphorylates and activates MAPK-K.

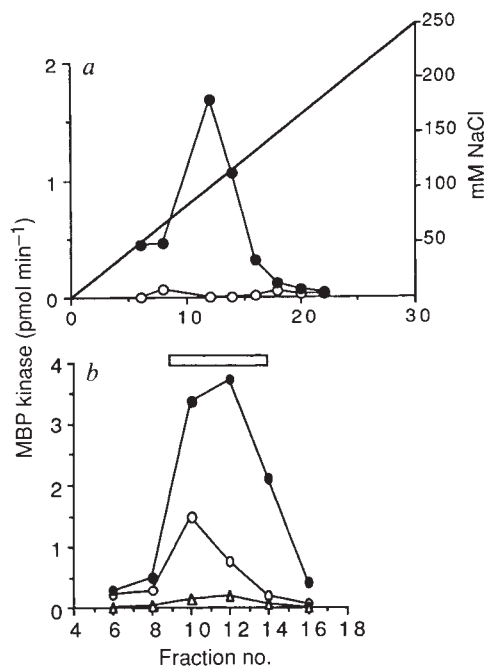
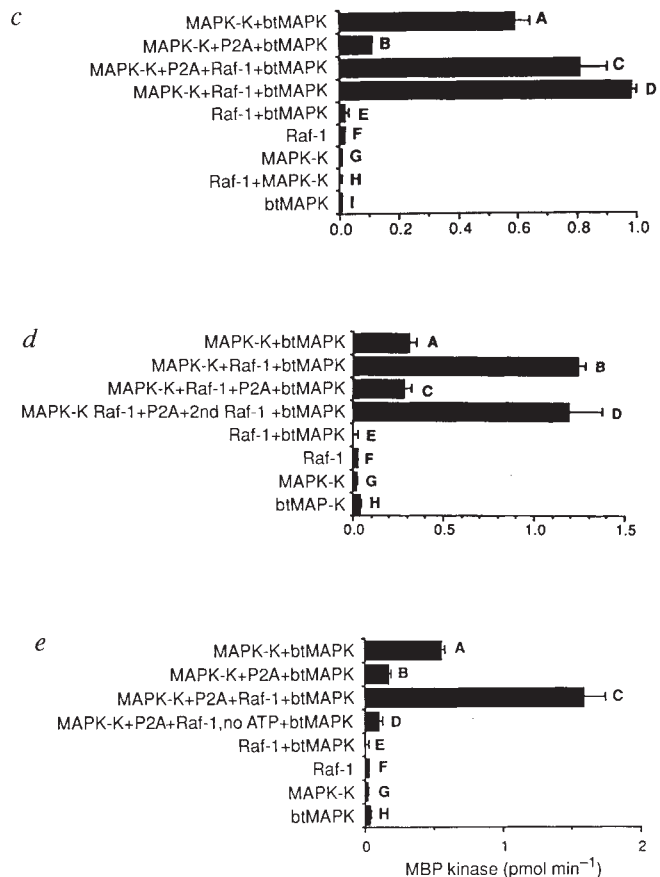


FIG. 3 Activation of MAPK-K by c-Raf-1. *a*, Mono-S chromatography (1-ml fractions) of insulin/vanadate-activated MAPK-K from H4 hepatoma cells. ●, Assayed for MBP kinase with added bacterial MAP kinase (btMAPK); ○, assayed for MBP kinase without added btMAP kinase (that is, endogenous MBP kinase, latent plus active). Basal btMAP kinase had an activity of 0.04 pmol min⁻¹. *b*, c-Raf-1 reactivates phosphatase 2A-deactivated MAPK-K. Mono-S fractions from *a* were assayed for MAPK-K directly (○), after phosphatase 2A treatment (△) or after phosphatase inactivation and subsequent incubation with Mg-ATP and c-Raf-1 (●). White bar indicates fractions pooled for further study (*c-f*). *c*, c-Raf-1 activates MAPK-K (with or without prior phosphatase 2A (P2A) treatment) beyond the initial level of MAPK-K activity. See text for discussion on bars A-I. *d*, MAPK-K-activated by c-Raf-1 *in vitro* is susceptible to deactivation by phosphatase 2A and reactivation by c-Raf-1, like the MAPK-K activated by insulin/vanadate *in situ*. *e*, c-Raf-1 activation of MAPK-K depends on Mg-ATP. For discussion on bars A-H, see text. *f*, c-Raf-1 phosphorylates a 50K polypeptide that coelutes from Mono-S with c-Raf-1-activated MAPK-K activity. Fraction numbers (6-16) are indicated along the top of the gel. *g*, Phosphoamino-acid analysis of the ³²P-labelled 50K species from *f*.

METHODS. H4 hepatoma cells (2.5×10^9 cells) were preincubated with VO₄ (0.2 mM) for 2 h and then treated with insulin (150 mU ml⁻¹) for 5 min. Cells were extracted by homogenization²⁶. MAPK-K was purified by ammonium sulphate precipitation (40-80% pellet), passing through DEAE-cellulose, followed by successive chromatography on Fast S-Sepharose (pH 6.5), DEAE-Cibacron blue 3GA agarose (eluted with ATP/NaCl), and Mono-S HR5/5 (pH 7.2) columns. The detailed procedure will be presented elsewhere (J.M.K. and J.A., manuscript in preparation). The specific activity of MAPK-K in the pooled Mono-S peak was 3,333 U mg⁻¹ (see text for unit definition) using the assay described in Fig. 1. Raf-1 was immunopurified from NIH BXB cells as follows. Twenty 10-cm plates were lysed into 2 ml lysis buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 50 mM β-glycerophosphate, 1 mM Na₃VO₄, 1% (w/v) Triton X-100, 10% (v/v) glycerol, 1 mM PMSF, 2 μM pepstatin A, 2 μM leupeptin, 10 kallekrein-inhibiting units per ml aprotinin). Lysates were cleared by centrifugation, and the supernatants incubated for 3 h at 4 °C with anti-Raf-1 peptide antibody SP63 or normal serum (1:200 dilution) plus 2-500 μl settled protein G-Sepharose. The protein G-Sepharose beads (immobilized Raf-1) were washed three times with lysis buffer (2 ml/wash), three times with LiCl wash (0.5 M LiCl, 100 mM Tris-HCl, pH 7.6), and twice



with either assay buffer (*a-c*, *f*, *g*) (50 mM β-glycerophosphate, pH 7.3, 1.5 mM EGTA, 1 mM DTT, 0.03% (w/v) Brij-35) or PIPES buffer (*d*) (10 mM PIPES, pH 7.2, 1 mM EGTA, 1 mM DTT, 0.03% (w/v) Brij 35, 5% (v/v) glycerol). Beads were divided among assay tubes. For MAPK-K (re)activation, H4 cell MAPK-K was treated with vehicle (PIPES buffer) or 5 U ml⁻¹ (ref. 24) purified phosphatase 2A for 35 min at 30 °C. Phosphatase was then inhibited with okadaic acid (5 μM). To 30 μl of treated MAPK-K, 30 μl 1:1 suspension of Raf-1 beads or normal serum beads were added along with 15 μl 50 mM MgCl₂/500 μM [γ-³²P]ATP (2-8,000 c.p.m. pmol⁻¹). Samples were incubated for 30 min at 30 °C with occasional vortexing. After 30 min, tubes were spun in a microfuge (4 °C) and 30 μl of the supernatant removed to a separate tube. To 30 μl of the supernatant were added 3 μl 1.4 mg ml⁻¹ btMAP kinase with 6 μl Mg-³²P ATP. Samples were incubated for a further 30 min at 30 °C and then 15 μl MBP reaction mix were added (150 mM Tris-HCl, pH 7.5, 3 mM DTT, 150 μM [γ-³²P]ATP (2-8,000 c.p.m. pmol⁻¹), 30 mM MgCl₂, 750 μg ml⁻¹ MBP, 15 μM heat-stable inhibitor of the cAMP-dependent protein kinase²¹). The reaction was continued for a further 15 min, when samples were quenched with SDS-EDTA and analysed by SDS-PAGE. MBP bands were excised from the stained, dried gels and counted by liquid scintillation spectroscopy. As controls, MBP kinase and MAPK-K contaminating the Raf-1 were determined, respectively, by incubating the Raf-1 with MBP mix alone (bar F in *c*, *d*, *e*) or with btMAP kinase followed by MBP mix (bar E in *c*, *d*, *e*). In addition, MBP kinase present in the MAPK-K preparation was measured (bar G in *c*, *d*, *e*), as was the basal MBP kinase activity present in the btMAP kinase (*c*, bar I; *d*, *e*, bar H). Data in *a-e* reflecting (re)activation of MAPK-K were corrected for these contaminating activities, which are shown in *c*, *d* and *e*, bars E-I, for comparison. All assays containing immunoprecipitated Raf-1 were also corrected for activity in the immunoprecipitates prepared with normal serum, which was routinely <2% of that seen in anti-Raf-1 precipitates. Bar graphs indicate mean ± s.d. for triplicate measurements. For the experiment in *f*, ³²P-labelled supernatant taken directly after incubation of Mono-S fractions with Mg²⁺ plus [γ-³²P]ATP, together with protein G-Sepharose beads prepared with either normal serum (left) or anti-Raf-1 antibody (right), were analysed by SDS-PAGE and autoradiography. For *g*, the ³²P-labelled 50K polypeptide seen in *f* was eluted, partially acid-hydrolysed (6 N HCl at 100 °C for 2 h) and analysed for phosphoamino acids by two-dimensional electrophoresis.

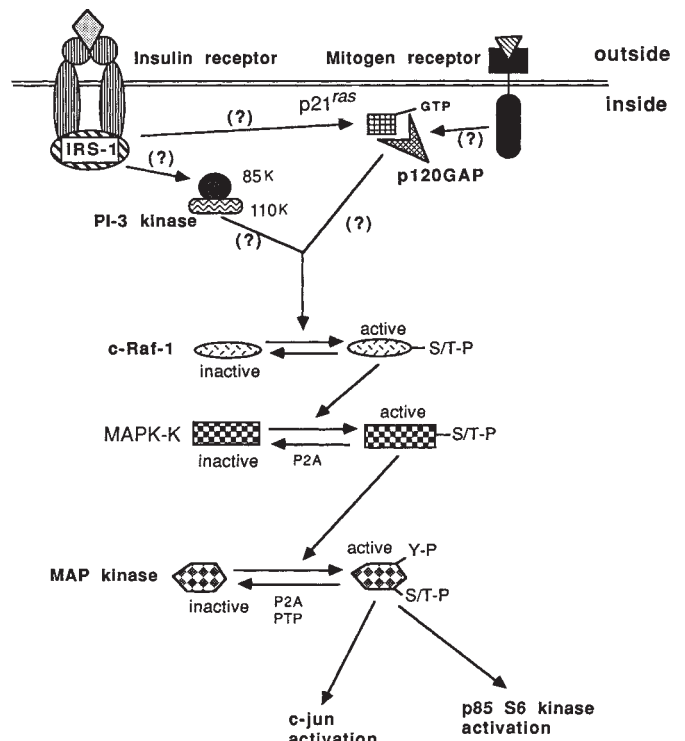
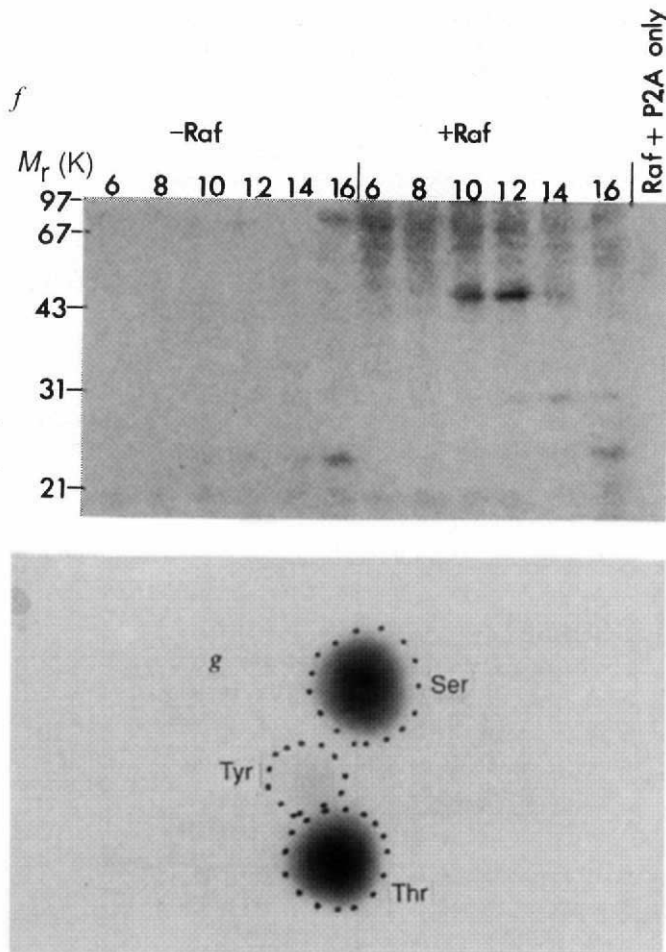


FIG. 4 Hypothetical model for an insulin/mitogen-stimulated protein phosphorylation cascade. For discussion, see text. Abbreviations: IRS-1, insulin receptor substrate-1; GAP, GTPase-activating protein; PI-3 kinase, phosphatidylinositol-3 kinase; S/T-P or Y-P, phosphorylated serine/threonine or tyrosine; P2A, phosphatase 2A; PTP, tyrosine phosphatase.

On the basis of these results, we propose a model for mitogen activation of protein phosphorylation (Fig. 4). Insulin- or mitogen-activated c-Raf-1 phosphorylates and activates MAPK-K, which in turn phosphorylates and activates Erk1 and 2. The latter in turn phosphorylates and activates c-Jun¹⁰ and the 85K (Rsk) S6 kinases^{5,11}. In skeletal muscle, insulin-activated p85 (Rsk) S6 kinase phosphorylates the G subunit of protein phosphatase-1, stimulating the activity of the phosphatase-1 catalytic subunit bound to glycogen granules through this G subunit. This reaction is proposed to underlie the insulin-induced dephosphorylation and activation of glycogen synthase¹².

The model shown in Fig. 4 portrays a different relationship between MAP kinases and c-Raf-1 from one proposed earlier^{13,14}, in which c-Raf-1 was inferred to be primarily downstream of the MAP kinases in the signal transduction cascade because c-Raf-1 is a substrate for MAP kinase *in vitro*. Those studies, however, did not observe activation of c-Raf-1 kinase catalytic activity after phosphorylation by MAP kinases, nor did MAP kinases phosphorylate *in vitro* all of the sites on c-Raf-1 phosphorylated *in situ* in response to mitogen¹³. Our results demonstrate directly that MAPK-K is activated by phosphorylation with c-Raf-1 kinase both *in vitro* and probably *in vivo*. From our results, MAP kinase phosphorylation of c-Raf-1 is unlikely to contribute to Raf-1 activation, but may represent a downregulatory modification.

It is not known what lies upstream of Raf-1. Logical candidates include phosphatidylinositol-3 kinase, which may generate ligand activators, and p21^{ras}/GTPase-activating protein. Mitogen and phorbol ester stimulation result in the appearance of the GTP-bound active form of p21^{ras} (refs 15–17). In addition, scrape loading of cells with oncogenic p21^{ras} activates MAP kinases *in vivo*¹⁸. Moreover, dominant negative raf mutants

inhibit p21^{ras}-mediated signalling¹⁹, and dominant negative ras mutants inhibit activation of c-Raf-1 by mitogens²⁰. The ras-related upstream elements may act directly on Raf-1 or through further intermediates such as a Raf-1 kinase-kinase. □

Received 30 March; accepted 8 June 1992.

- Rapp, U. R. *Oncogene* **6**, 495–500 (1991).
- Ahn, N. G., Weiel, J. E., Chan, C. P. & Krebs, E. G. *J. Biol. Chem.* **265**, 11487–11494 (1990).
- Kyriakis, J. M., Brautigan, D. L., Ingelbritsen, T. S. & Avruch, J. *J. Biol. Chem.* **266**, 10043–10046 (1991).
- Anderson, N. G., Maller, J. L., Tonks, N. K. & Sturgill, T. W. *Nature* **343**, 651–653 (1990).
- Ahn, N. G., et al. *J. Biol. Chem.* **266**, 4220–4227 (1991).
- Gomez, N. & Cohen, P. *Nature* **353**, 170–173 (1991).
- Alleman, G. L., Her, J.-H., Wu, J., Sturgill, T. W. & Weber, M. *J. Molec. cell. Biol.* **12**, 2222–2229 (1992).
- Rapp, U. R., Reynolds, F. H. & Stephenson, J. R. *J. Virol.* **45**, 914–924 (1983).
- Matsuda, S. et al. *EMBO J.* **11**, 973–982 (1992).
- Pulverer, B. J., et al. *Nature* **353**, 670–674 (1991).
- Sturgill, T. W., Ray, L. B., Erikson, E. & Maller, J. L. *Nature* **334**, 715–718 (1988).
- Dent, P. et al. *Nature* **348**, 302–308 (1990).
- Anderson, N. G. et al. *Biochem. J.* **277**, 573–576 (1991).
- Lee, R.-M., Cobb, M. H. & Blakeshear, P. *J. J. Biol. Chem.* **267**, 1088–1092 (1992).
- Downward, J. et al. *Nature* **346**, 719–723 (1990).
- Burgering, B. M. et al. *EMBO J.* **10**, 1103–1109 (1991).
- Satoh, T., Nakafuku, M., Miyajima, A. & Kaziro, Y. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3314–3318 (1991).
- Leever, S. J. & Marshall, C. J. *EMBO J.* **11**, 569–574 (1992).
- Bruder, J. T., Heidecker, G. & Rapp, U. R. *Genes Dev.* **6**, 545–556 (1992).
- Troppair, J. et al. *Oncogene* (in the press).
- Grove, J. R., Price, D. J., Goodman, H. M. & Avruch, J. *Science* **238**, 530–533 (1987).
- Hsi, E. D. et al. *J. Biol. Chem.* **264**, 10836–10842 (1989).
- Kyriakis, J. M. & Avruch, J. *J. Biol. Chem.* **265**, 17355–17363 (1990).
- Takeda, M. et al. *J. Biol. Chem.* **258**, 10455–10463 (1988).
- Jones, S. W., Erikson, R. L., Ingelbritsen, V. M. & Ingelbritsen, T. S. *J. Biol. Chem.* **264**, 7747–7753 (1989).
- Kyriakis, J. M. & Avruch, J. *Biochem. Biophys. Acta* **1054**, 73–82 (1990).

ACKNOWLEDGEMENTS. We thank T. Ingelbritsen for rrbPTP-1; T. Force and J. Bonventre for discussion; D. Brady for technical assistance; M. Chambers for help in preparation of the manuscript; and S. Bennett, K. Grady and M. Byrne for photography. This work was supported by a US Public Health Service FIRST Award to J.M.K. and grants from the American Cancer Society, the United States Public Health Service and a Mentor-based Fellowship from the American Diabetes Association (supporting X.-F.Z.) to J.A.

Specific binding of the transcription factor sigma-54 to promoter DNA

Martin Buck & Wendy Cannon

AFRC Nitrogen Fixation Laboratory, University of Sussex, Brighton BN1 9RQ, UK

A CENTRAL event in transcription is the assembly on DNA of specific complexes near the initiation sites for RNA synthesis. Activation of transcription by one class of enhancer-binding proteins requires an RNA polymerase holoenzyme¹ containing the specialized transcription factor, sigma-54 (σ^{54}). We report here that σ^{54} alone specifically binds to promoter DNA and is responsible for many of the close contacts between RNA polymerase holoenzyme and promoter DNA, a property proposed for the major σ^{70} protein family. Binding of σ^{54} to promoter DNA is not equivalent to that of holoenzyme suggesting that there is a constraint on σ^{54} conformation when bound with core RNA polymerase. Footprints indicate σ^{54} is at the leading edge of DNA-bound holoenzyme. Like the holoenzyme¹⁻⁴, σ^{54} -binding to promoter DNA does not result in DNA strand separation. Instead the specific DNA-binding activity of σ^{54} assists assembly of a closed promoter complex. This complex can be isomerized to the open (DNA melted) complex by activator protein^{5,6}, but promoter-bound σ^{54} alone cannot be induced to melt DNA. The pathway leading to productive transcription is similar to that proposed for eukaryotic RNA polymerase II systems.

The promoter sequence recognized by the holoenzyme ($E\sigma^{54}$) is generally characterized by the presence of GG and GC doublets 24 and 12 base pairs, respectively, upstream of the transcription initiation point (Table 1; refs 1, 3), and activation requires binding of the appropriate activator protein upstream of $E\sigma^{54}$ at enhancer sequences⁷⁻⁹. Using conditions for detecting closed promoter complexes at the σ^{54} -dependent *Klebsiella pneumoniae* nitrogen fixation (*nif*) promoters¹⁰ we footprinted the *nifH* promoter and a mutant variant, *nifH049* (Table 1) which forms a stronger closed complex^{3,11}. We were able to detect $E\sigma^{54}$ complexes bound at the *nifH* and *nifH049* promoters, and σ^{54} bound at the *nifH049* promoter (Figs 1a and 3). The *nifH* promoter did not detectably bind σ^{54} (Fig. 3, lanes 16, 17). The complexes with $E\sigma^{54}$ and the *nifH* and *nifH049* promoters were not equivalent. The *ExoIII* digestion block was 3–4 base pairs (bp) closer to the transcription initiation site in the *nifH049* promoter. Thus $E\sigma^{54}$ seemed to establish more contacts at the *nifH049* promoter. In the absence of core RNA polymerase subunits, σ^{54} bound to the *nifH049* promoter to produce a clear *ExoIII* block (Fig. 1a), but only partially protected conserved guanine residues at -24, -25 and -13 from dimethylsulphate attack (Fig. 1b). The σ^{54} -dependent *ExoIII* block at the *nifH049* promoter was 3–4 bp closer to the transcription initiation site than when $E\sigma^{54}$ bound (Fig. 1a), indicating σ^{54} is conformationally constrained by core polymerase. Unlike $E\sigma^{54}$ (ref. 11), binding of σ^{54} did not increase reactivity of the T at -9 towards $KMnO_4$ (data not shown). Thus σ^{54} did not seem to distort the DNA. DNase 1 footprints revealed that the DNA was covered to a similar extent by σ^{54} and $E\sigma^{54}$, but that the major $E\sigma^{54}$ footprint extended about 7 bases further upstream than did the σ^{54} footprint (Fig. 1c). Gel mobility shift assays also demonstrated that the complex forming with σ^{54} was different to that forming in the presence of $E\sigma^{54}$, the latter complex having a much reduced mobility (data not shown).

To extend our analysis of σ^{54} binding to include a naturally occurring promoter sequence (Table 1) we examined the binding of $E\sigma^{54}$ (ref. 3) and σ^{54} to the *Rhizobium meliloti nifH* promoter (Fig. 2). Both σ^{54} and $E\sigma^{54}$ produced clear blocks to *ExoIII* digestion, with the σ^{54} block being about 4 bp closer to the

transcription initiation site than the $E\sigma^{54}$ block. Dimethylsulphate and DNase 1 footprints again revealed that σ^{54} was responsible for most of the close contacts detected between holoenzyme and the *R. meliloti nifH* promoter (Gs at -14, -25, -26; Fig. 2). Holoenzyme covered ~32 bp and σ^{54} covered 27 bp of promoter DNA as judged by DNase 1 footprinting (Fig. 2). Footprints on the bottom strand showed a single *ExoIII* block for σ^{54} and $E\sigma^{54}$ at around -34, protection of Gs -13, -22 and -24 by σ^{54} and $E\sigma^{54}$ and a DNase 1 footprint from -34 to -5 for σ^{54} and $E\sigma^{54}$.

The *nifH049* and *R. meliloti nifH* promoters have a 'T'-tract from -14 to -17 that distinguishes them from the *nifH* promoter¹¹ and is necessary for the binding of σ^{54} (Table 1). *R. meliloti* and *nifH049* templates in which dT was replaced by dU, effectively removing the 5-methyl group of T, bound $E\sigma^{54}$ but not σ^{54} (Fig. 3; Table 1). In complementary experiments we constructed wild-type sequence *K. pneumoniae nifH* templates in which dC was replaced by dm⁵C and found that σ^{54} then bound (Fig. 3). The sequence from -14 to -17 is therefore critical for σ^{54} binding. Specifically the methyl groups in the DNA major groove are important (Table 1).

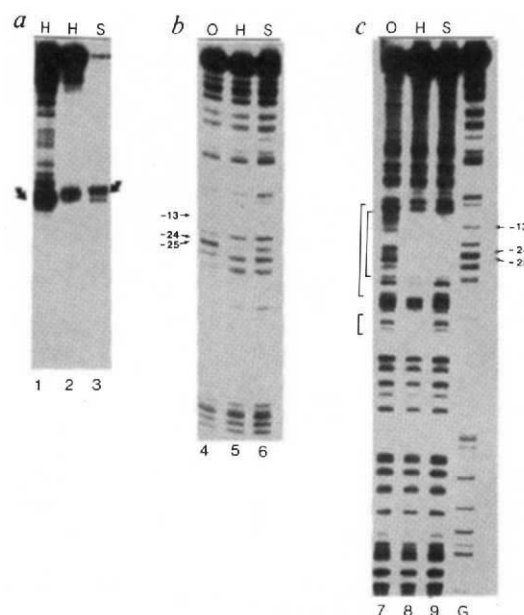


FIG. 1 Footprints demonstrating the binding of $E\sigma^{54}$ and σ^{54} to the *K. pneumoniae nifH* and *nifH049* promoters. a, Exonuclease III footprints; b, dimethylsulphate footprints; c, DNase 1 footprints; H, Holoenzyme; S, σ^{54} ; O, no addition. Arrows mark *ExoIII* blocks due to $E\sigma^{54}$ or σ^{54} binding; brackets, the extent of the DNase 1 footprint. Lane 1, wild-type *nifH* promoter; lanes 2–9, *nifH049* promoter. Lane G, Chemical cleavage of DNA at G residues. $E\sigma^{54}$ bound the *nifH* and *nifH049* promoters, σ^{54} bound the *nifH049* promoter. Footprints were conducted as described previously¹⁰ in 25 mM Tris-acetate, 8 mM magnesium acetate, 10 mM KCl, 1 mM dithiothreitol, 3.5% (w/v) polyethylene glycol, pH 8.0. Template DNA (2.4 nM) was incubated with 125 nM core RNA polymerase (E) and 280 nM σ^{54} ($E\sigma^{54}$) (E and σ^{54} were previously combined and warmed at 30 °C for 5 min before addition) at 30 °C in a 50- μ l assay. When σ^{54} (560 nM) alone was used, this was also preincubated before adding to template DNA. Binding was for 20 min before addition of footprinting reagent. In DNase 1 footprints, salmon sperm DNA (270 μ g ml⁻¹) was included in the incubations. *K. pneumoniae* σ^{54} was free of β and β' RNA polymerase subunits as judged by stained, overloaded gels. Equivalent exonuclease III and DMS footprints were obtained when the binding buffer was supplemented with 100 mM KCl or footprints were conducted in 50 mM Tris-acetate, 100 mM potassium acetate, 8 mM magnesium acetate, 27 mM ammonium acetate, 1 mM dithiothreitol, 3.5% (w/v) polyethylene glycol². Thus binding of σ^{54} persists at physiologically relevant salt concentrations. Positions of digestion blocks were estimated from G and C+T chemical cleavage sequencing ladders.

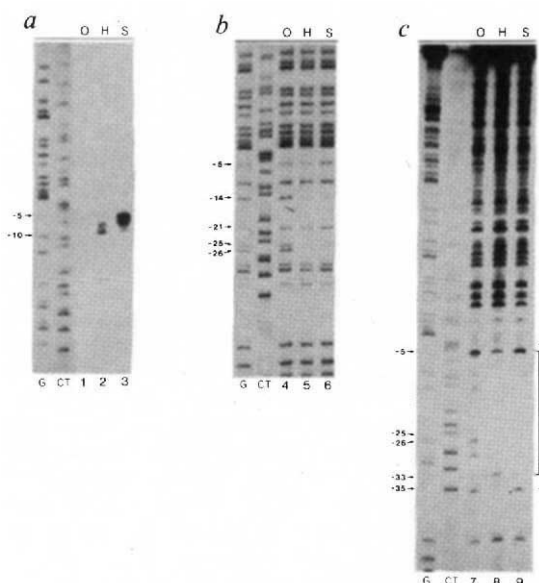


FIG. 2 Footprints demonstrating the binding of $E\sigma^{54}$ (H) and σ^{54} (S) to the *R. meliloti nifH* promoter. *a*, Exonuclease III footprints; *b*, dimethylsulphate footprints; *c*, DNase I footprints. Conserved Gs at -14, -25, -26 were protected from dimethylsulphate attack by σ^{54} and $E\sigma^{54}$. Assays were as for Fig. 1, and footprints persisted in buffer supplemented with 100 mM KCl or when the 100 mM potassium acetate-containing buffer was used. The σ^{54} *ExoIII* block persisted down to 15 nM σ^{54} , and shifting the σ^{54} block to that characteristic of $E\sigma^{54}$ needed a minimum of 2 nM E at 280 nM σ^{54} . Serial dilution of $E\sigma^{54}$ resulted in a gradual loss of the $E\sigma^{54}$ *ExoIII* block without the appearance of the block characteristic of σ^{54} alone.

The closed complex that forms at the *nifH049* promoter can be isomerized to the open complex by the NtrC S160F enhancer-binding activator protein¹¹⁻¹³. But when prebound σ^{54} at *nifH049* was incubated with NtrC S160F protein no evidence for DNA melting (using DNA reactivity towards $KMnO_4$ as an index of denaturation) was obtained (data not shown). Thus σ^{54} seems unable to function in the process of open complex formation unless bound as holoenzyme.

In summary, σ^{54} functions through assisting assembly of a preinitiation complex consisting of σ^{54} -holoenzyme and promoter DNA at a site centred about 18 bp upstream of the

TABLE 1 Promoter sequences and template requirements for binding σ^{54}

Promoters		σ^{54} Binding
Wild-type <i>nifH</i>	CTGGTATGTTCCCTGCA	—
<i>nifH049</i>	CTGGTATGTTTTTGCA	+
<i>R. meliloti nifH</i>	CTGGACGACTTTTGCA	+

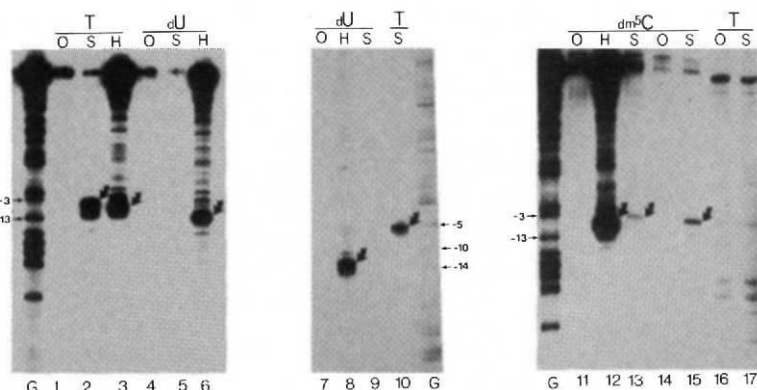
Template sequences	$E\sigma^{54}$ Footprints	σ^{54}
<i>nifH</i>	CCCTGCA	+++
<i>nifH049</i>	TTTTGCA	+++
<i>RmH</i>	TTTTGCA	+++
<i>nifH049</i> -dU	UUUUGCA	+++
<i>RmH</i> -dU	UUUUGCA	+++
<i>nifH</i> -dm ⁵ C	m ⁵ Cm ⁵ Cm ⁵ CTGm ⁵ CA	+++

Binding sites for $E\sigma^{54}$ are characterized by GG and GC doublets separated by 10 bp and centred roughly 18 bp upstream of the transcription initiation site. A T-tract upstream of the GC doublet is important for the binding of $E\sigma^{54}$ and σ^{54} (refs 3, 22). The *nifH049* and *R. meliloti nifH* promoters bound σ^{54} but binding to the wild-type *nifH* promoter was not detected (Figs 1–3). When templates were prepared in which the 5-methyl group of T was absent (dU), σ^{54} did not detectably bind (Fig. 3). Introduction of 5-methyl (dm⁵C) into templates increased binding of σ^{54} (Fig. 3).

transcription initiation point. A predominant event in binding holoenzyme to promoter DNA is recognition of DNA by σ^{54} . The nonequivalence in the σ^{54} recognition sequence when the -25 and -13 elements are compared presumably enables the orientation of DNA-bound σ^{54} to define the direction of transcription.

Detection of $E\sigma^{54}$ binding when σ^{54} binding is not evident indicates core RNA polymerase reduces the promoter dissociation rate of σ^{54} . It is possible that the binding specificities of σ^{54} and $E\sigma^{54}$ are different, suggesting a source of favourable binding energy for the holoenzyme. Notably σ^{54} functions with the same core RNA polymerase subunits as does the major eubacterial sigma factor, σ^{70} , but seems to associate (at least in the absence of DNA template) relatively weakly with core RNA polymerase¹⁴. It is possible that σ^{54} can bind the template and then associate with core RNA polymerase, possibly in contrast to σ^{70} which seems to bind only as holoenzyme¹⁵. It seems that σ^{54} is at the leading edge of the holoenzyme, positioned to extend over DNA to be melted during open complex formation, implying σ^{54} might directly participate in DNA melting. Open promoter complex formation occurs efficiently in response to

FIG. 3 Binding of $E\sigma^{54}$ (H) and σ^{54} (S) to templates altered by the replacement of dT by dU or of dC by dm⁵C (see also Table 1). Binding was assayed by exonuclease III footprinting and blocks are marked with arrows. Lanes 1–3, *nifH049* dT template footprints; lanes 4–6, *nifH049* dU template footprints; lanes 7–9, *R. meliloti nifH* dU template footprints; lane 10, *R. meliloti nifH* dT template footprint. Wild-type *K. pneumoniae nifH* dm⁵C template footprints, lanes 11–15. For comparison, footprints of the wild-type *K. pneumoniae* dT template are shown in lanes 16 and 17. Assays for lanes 15 and 17 contained 1.2 μ M σ^{54} . Lanes 16 and 17 were twofold overloaded. Lane G, chemical cleavage of DNA at G residues. Templates were generated by polymerization using nucleotide triphosphate mixes containing either dUTP or dm⁵CTP rather than dTTP or dCTP, and thus the top non-template strand only is modified (the strand visualized in the footprints). Substituting dT by dU effectively removes the C⁵-methyl group of dT found in the major groove of DNA, and replacement of dC by dm⁵C effectively introduces C⁵-methyl. No binding of σ^{54} to the wild-type *nifH* dT template was detected (lane 17 and unpublished data). Replacement of dT by dU greatly reduced σ^{54} binding (lanes 5, 9). Replacement of dC by dm⁵C increased σ^{54} binding (lanes 13, 15). Differences in DNA geometries may also contribute to the preferential binding of σ^{54} to dT rather than dm⁵C containing templates. As anticipated from the conservation of G-C base



pairs at -24, -25 and -13 and their protection from methylation by $E\sigma^{54}$ and σ^{54} (Fig. 2; refs 2, 10) replacement of dG by C⁷-deaza guanine resulted in templates that failed to bind $E\sigma^{54}$ or σ^{54} as judged by DNase I footprinting (data not shown). The sequence TTGCT rather than TTTGCA from -16 to -11 in the enteric *glnAp2* promoters may explain why the binding of σ^{54} is not readily detected at this sequence^{19,20}. Proposed domain structures for σ^{54} indicate that DNA recognition may involve a helix-turn-helix structure^{18,21}.

the activator only when σ^{54} is bound as holoenzyme at the promoter. The conformation of σ^{54} in the holoenzyme, or the RNA polymerase subunits themselves, may also be important for DNA melting. That σ^{54} binds to promoter DNA but does not melt it is consistent with all known σ^{54} -dependent transcription systems that have a strict requirement for an activator protein¹. Separation of promoter recognition and DNA melting functions permits $E\sigma^{54}$ to respond to many stimuli of gene expression through their influences on the activity of the activator protein. Activator dependence and assembly of the preinitiation complex in its absence strongly suggests a broad similarity to eukaryotic RNA polymerase II transcription^{16,17}. □

Received 24 March; accepted 10 June 1992.

1. Kustu, S., Santero, E., Keener, J., Popham, D. & Weiss, D. *Microbiol. Rev.* **53**, 367–379 (1989).
2. Popham, D. L., Szeto, D., Keener, J. & Kustu, S. *Science* **243**, 629–635 (1989).
3. Morett, E. & Buck, M. *J. molec. Biol.* **210**, 65–77 (1989).
4. Sasse-Dwight, S. & Gralla, J. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8934–8938 (1988).
5. Weiss, D. S., Batut, J., Klose, K. E., Keener, J. & Kustu, S. *Cell* **67**, 155–167 (1991).
6. Austin, S. & Dixon, R. *EMBO J.* **6**, 2219–2228 (1992).
7. Buck, M., Miller, S., Drummond, M. & Dixon, R. *Nature* **320**, 374–378 (1986).
8. Reitzer, L. J. & Magasanik, B. *Cell* **45**, 785–792 (1986).
9. Morett, E. & Buck, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9401–9405 (1988).
10. Buck, M. & Cannon, W. *Molec. Microbiol.* (in the press).
11. Cannon, W. & Buck, M. *J. molec. Biol.* **225**, 271–286 (1992).
12. Popham, D., Kenner, J. & Kustu, S. *J. biol. Chem.* **266**, 19510–19518 (1991).
13. Austin, S., Kundrot, C. & Dixon, R. *Nucleic Acids Res.* **19**, 2281–2287 (1991).
14. Lesley, S. A., Brow, M. A. D. & Burgess, R. *J. biol. Chem.* **266**, 2632–2638 (1991).
15. Helman, J. D. & Chamberlain, M. J. *A. Rev. Biochem.* **57**, 839–872 (1988).
16. Greenblatt, J. *Cell* **66**, 1067–1070 (1991).
17. Wang, W., Carey, M. & Gralla, J. D. *Science* **255**, 450–453 (1992).
18. Sasse-Dwight, S. & Gralla, J. D. *Cell* **62**, 945–954 (1990).
19. Kustu, S., Sei, K. & Keener, J. in *Regulation of Gene Expression, Symp. Soc. gen. Microbiol.* (eds Booth, I. & Higgins, C.) 139–154 (Cambridge Univ. Press, 1986).
20. Ninfa, A. J., Reitzer, L. J. & Magasanik, B. *Cell* **50**, 1039–1046 (1987).
21. Coppard, J. R. & Merrick, M. J. *Molec. Microbiol.* **5**, 1309–1317 (1991).
22. Claverie-Martin, F. Magasanik, B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1631–1635 (1991).

ACKNOWLEDGEMENTS. We thank S. Austin for purified *K. pneumoniae* proteins; M. Moore for discussions; S. Austin, R. Dixon, M. Drummond, M. Merrick and B. Smith for comments on the manuscript and R. Foote for typing.

Different β -subunits determine G-protein interaction with transmembrane receptors

C. Kleuss*, H. Scherübl†, J. Hescheler†, G. Schultz† & B. Wittig*

* Institut für Molekularbiologie und Biochemie, and

† Institut für Pharmakologie, Freie Universität Berlin, Arnimallee 22, D-1000 Berlin 33, Germany

REGULATORY GTP-binding proteins (G proteins) are membrane-attached heterotrimers (α , β , γ) that mediate cellular responses to a wide variety of extracellular stimuli^{1,2}. They undergo a cycle of guanine-nucleotide exchange and GTP hydrolysis, during which they dissociate into α -subunit and $\beta\gamma$ complex¹. The roles of G-protein α -subunits in these processes and for the specificity of signal transduction are largely established; the β - and γ -subunits are essential for receptor-induced G-protein activation and seem to be less diverse and less specific. Although the complementary DNAs for several β -subunits have been cloned^{2,5–8}, isolated subunits have only been studied as $\beta\gamma$ complexes^{3,9–12}. Functional differences have been ascribed to the γ -subunit on the basis of extensive sequence similarity among β -subunits and apparent heterogeneity in γ -subunit sequences^{13,14}. $\beta\gamma$ complexes can interact directly or indirectly with different effectors^{10,11,15–20}. They seem to be interchangeable in their interaction with pertussis toxin-sensitive α -subunits³, so we tested this by microinjecting antisense oligonucleotides into nuclei of a rat pituitary cell line to suppress the synthesis of individual β -subunits selectively. Here we show that two out of four subtypes of β -subunits tested (β_1

and β_3) are selectively involved in the signal transduction cascades from muscarinic M_4 (ref. 4) and somatostatin receptors, respectively, to voltage-dependent Ca^{2+} channels.

We have established nuclear microinjection of short selective antisense oligonucleotides as a general method to study G protein function in intact cells²¹. The effects of 'knocking out' the expression of individual G-protein α -subunits are measured in a single cell either electrophysiologically or by immunofluorescent labelling of the targeted proteins. Here we apply this approach to elucidate the role of β -subunit subtypes in the differential coupling of G proteins to the same receptor/G protein/effector systems in GH₃ cells, a rat pituitary cell line.

Cells that have been injected with the antisense oligonucleotide β -com (see Fig. 1 legend), which targets messenger RNA of all known four β -polypeptides, contain reduced amounts of immunostainable β -subunits (data not shown). The disappearance of β -subunits parallels the functional loss of inhibitory hormonal effects on Ca^{2+} currents. Ca^{2+} currents in GH₃ cells are probably due to the activation of L-type channels and are inhibited to 70–80% of control currents by carbachol and somatostatin²¹. One day after injection of the oligonucleotide β -com, neither hormone reduced the Ca^{2+} current (Fig. 1); this effect lasted for one more day. The original hormone sensitivity of the Ca^{2+} channel was restored by ~60 h after injection.

To study the question of preferential interaction between the activated receptor and G_O proteins containing a given β -subtype, GH₃ cells were injected with antisense oligonucleotides that can specifically hybridize with the mRNA of one particular β -subtype (Figs 2 and 3). Neither β_2 - nor β_4 -specific antisense

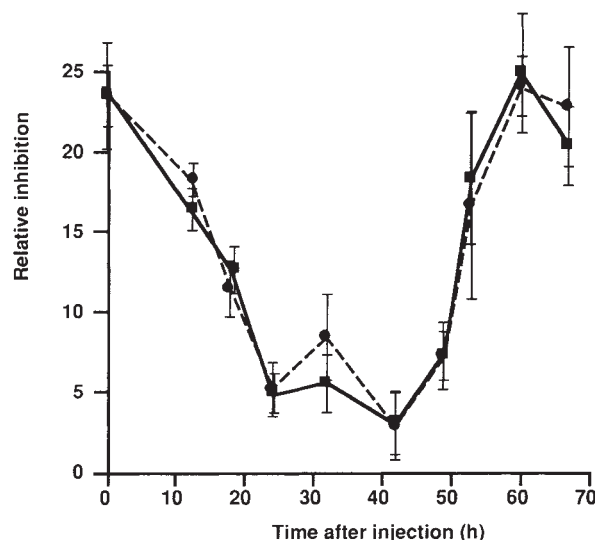
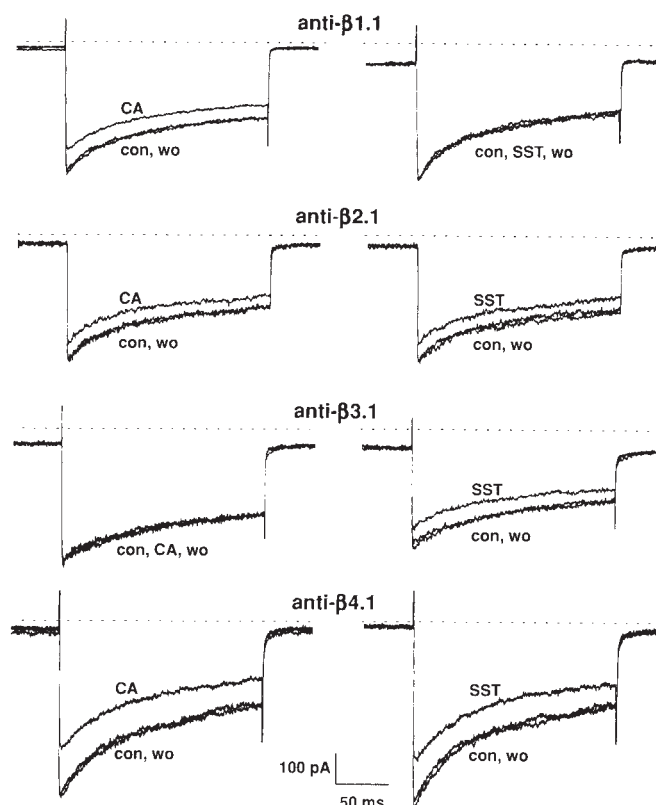


FIG. 1 Time course of Ca^{2+} current inhibition by somatostatin or carbachol in GH₃ cells. At time zero, cells were injected with antisense oligonucleotide β -com (see Methods). At the indicated time points, the Ca^{2+} current was measured in the presence of 1 μ M somatostatin (dotted line) or 10 μ M carbachol (solid line). Mean values with s.e.m. are shown ($n \geq 5$).

METHODS. Microinjection (about 10,000 full-length oligonucleotides per nucleus) and electrophysiological measurements were done as described²¹, but during current recording cells were perfused with a solution containing 120 mM choline-Cl, 10.8 mM $BaCl_2$, 1 mM $MgCl_2$, 5.4 mM $CsCl$, 10 mM glucose and 10 mM tetraethylammonium-HEPES (pH 7.4 at 37 °C). Sequence of injected β -com oligonucleotide: 5'-TTGACGTTGAAGTCGTCRTA-3', corresponding to nucleotides 825–844 of the identical strand of the β_1 gene sequence⁵. It can hybridize with the mRNAs of β_1 , β_2 , β_3 and β_4 . Abbreviations for wobbled positions: R (G or A), M (A or C), Y (T or C), S (G or C). The oligonucleotide β -com is different from the previously used oligonucleotide anti- β ²¹. Oligonucleotide β -com can hybridize perfectly with β_1 , β_2 , and β_3 mRNAs, and the longest continuously matching stretch in β_4 mRNA is 16 bases. The much-improved hybridization characteristics compared with oligonucleotide anti- β may explain why less β -com (<10,000 molecules, as compared with 50,000 for anti- β) is needed for effective repression of hormone responses.

FIG. 2 Time-current recordings of the voltage-sensitive Ca^{2+} channels in GH_3 cells in the presence of carbachol (left-hand panels) or somatostatin (right-hand panels). Current traces are shown for one cell, each superfused with either hormone at 40 h after injection with antisense oligonucleotides anti- $\beta 1.1$, anti- $\beta 2.1$, anti- $\beta 3.1$ or anti- $\beta 4.1$. Under voltage-clamp conditions, whole-cell Ca^{2+} currents were recorded by depolarizing pulses from -80 to 0 mV. Abbreviations: con, control currents obtained before application of receptor agonist; SST, currents recorded during superfusion of cells with $1 \mu\text{M}$ somatostatin; CA, currents recorded during superfusion of cells with $10 \mu\text{M}$ carbachol; wo, currents recorded after removal of receptor agonists. Sequences of injected oligonucleotides: anti- $\beta 1.1$: 5'-GAGAGAGAGTTGC-ATCTGC-3', corresponding to nucleotides 75-93 of the identical strand of the β_1 gene sequence⁵; anti- $\beta 2.1$: 5'-GGGTCAGTGTGAGTCCCC-3', corresponding to nucleotides 76-94 of the identical strand of the β_2 gene sequence⁶; anti- $\beta 3.1$: 5'-GGCCAGACACCAGCTCTGCC-3', corresponding to nucleotides 90-109 of the identical strand of the β_3 gene sequence⁷; anti- $\beta 4.1$: 5'-GAACCAGCGTGGCATCGTT-3', corresponding to nucleotides 76-94 of the identical strand of the β_4 gene sequence⁸. Each can selectively hybridize with the described mRNA.



oligonucleotides prevented the inhibitory effects of somatostatin or carbachol on the Ca^{2+} channel. However, Ca^{2+} currents of cells that had been injected with antisense oligonucleotide anti- $\beta 1.1$ were no longer inhibited by somatostatin, whereas carbachol inhibited as it did in non-injected cells. This suggests that a G protein containing the β_1 subtype (running on a denaturing polyacrylamide gel with an apparent relative molecular mass of 36,000 couples to the somatostatin receptor. Data are consistent with results showing that the β_{36K} subunit

coprecipitates with the somatostatin receptor solubilized from rat brain²². In an analogous experiment, the Ca^{2+} current of cells that had been injected with antisense oligonucleotide anti- $\beta 3.1$ was no longer inhibited by carbachol, whereas somatostatin was still effective. The differential receptor coupling of G_O proteins with β_1 and β_3 subunits was confirmed by injection of two other specific antisense oligonucleotides, anti- $\beta 1.2$ and anti- $\beta 3.2$, respectively (Fig. 3). We conclude that in these signal transduction pathways one particular $\alpha\beta\gamma$ complex out of a

FIG. 3 Ca^{2+} current inhibition by receptor agonists in GH_3 cells injected with antisense oligonucleotides directed against mRNAs encoding β -subunit polypeptides. Whole-cell Ca^{2+} currents were measured about 40 h after injection of the oligonucleotide. Shown is the inhibition of Ca^{2+} currents by somatostatin ($1 \mu\text{M}$; hatched bars) or carbachol ($10 \mu\text{M}$; dark bars) in per cent of control current observed in the absence of the respective agonist (mean values with s.e.m. are shown). On the left-hand side are listed the numbers of cells that were used for electrophysiological measurements: A, cells successively treated with either receptor agonist (random order) with intermediate washing; B, cells treated with carbachol only; C, cells treated with somatostatin only; D, cells without measurable Ca^{2+} currents in the absence of receptor agonists; E, cells inhibited by carbachol by more than 10% of control currents; F, cells inhibited by somatostatin by more than 10% of control currents. For sequences of injected oligonucleotides βcom , anti- $\beta 1.1$, anti- $\beta 2.1$, anti- $\beta 3.1$ and anti- $\beta 4.1$, see legends to Figs 1 and 2. Sequence of anti- $\beta 1.2$: 5'-TGTCGGTAAACGTGGTCGTCT-3', corresponding to nucleotides 526-547 of the identical strand of the β_1 gene sequence⁵. It can selectively hybridize with the described β_1 mRNA; anti- $\beta 3.2$: 5'-ACGTGACGACAGGCTTTCCT-3', corresponding to nucleotides 64-83 of the identical strand of the β_3 gene sequence⁷. It can selectively hybridize with the described β_3 mRNA.

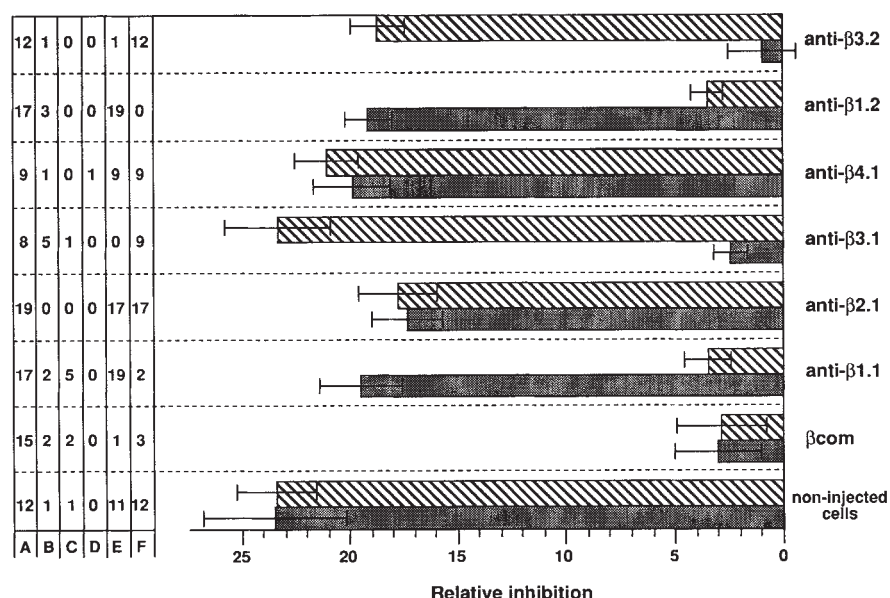
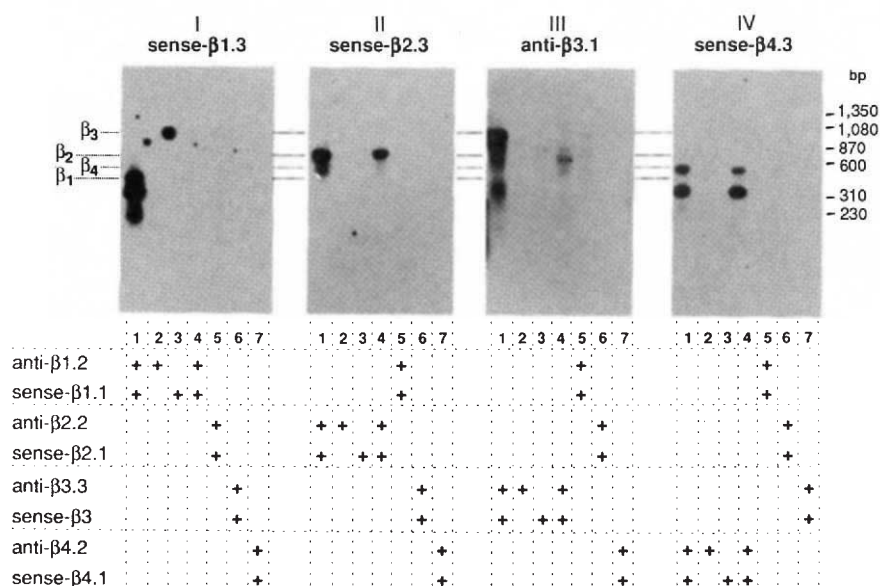


FIG. 4 Existence of mRNA encoding G protein β -subunits in GH₃ cells. mRNA (1.4 μ g) isolated from GH₃ cells was reverse-transcribed with 200 U of Moloney murine leukaemia virus reverse transcriptase (BRL) in 20 μ l of 0.5 mM each of dATP, dCTP, dGTP and dTTP, 1 μ g random primer (dN₆), 75 mM KCl, 10 mM dithiothreitol, 3 mM MgCl₂, 50 mM Tris-HCl, pH 8.3, for 1 h at 37 °C. cDNA within 1/100 part reaction volume served as template for an amplification reaction using oligonucleotides that hybridize with the β_1 , β_2 , β_3 and β_4 gene sequences. The polymerase chain reaction (PCR) was done in a 50- μ l reaction volume using 1 U of *Taq* polymerase (Promega). After 93 °C for 1.5 min, 50 °C for 10 s, 72 °C for 1 min (35 cycles) in a DNA Thermal Cycler (Perkin Elmer), reaction products were precipitated, electrophoretically separated on an agarose gel (1.5% w/v) and transferred to a nylon membrane (Gene Screen, NEN DuPont).

After crosslinking the nucleic acids to the membrane with ultraviolet light (120 mJ, 254 nm), [³²P]-labelled oligonucleotides were hybridized according to ref. 23. To achieve stringent hybridization conditions we varied the incubation temperature between 40 °C and 60 °C. Autoradiograms are shown of filters that had been hybridized with sense- β_1 3 (I), sense- β_2 3 (II), anti- β_3 1 (III), or sense- β_4 3 (IV). cDNA from GH₃ (lanes 1–3, 5–7) and RBL-2H3 cells (rat basophilic leukaemic cell line, lane 4) were amplified by PCR using primers as indicated under the panels. Sequences of oligonucleotides: for anti- β_1 2 and anti- β_3 1, see legends to Figs 2 and 3; sense- β_1 1, sense- β_1 3: 5'-GCAGATGCAACTCTCTC-3' and 5'-CTCCATTTCACATCTGAAACT-3', corresponding to nucleotides 75–93 and 362–383 of the identical strand of the β_1 gene sequence⁵, respectively; sense- β_2 1, sense- β_2 3: 5'-GGGACTCAACACTGACCC-3' and 5'-GACTCAAGGCTGTCGTCAGCG-3', corresponding to nucleotides 76–94 and 196–217 of the identical strand of the β_2 gene sequence⁶, respectively; sense- β_3 : 5'-AGATTGCAGATGCCAG-



GAAA-3', corresponding to nucleotides 70–89 of the identical strand of the β_3 gene sequence⁷; sense- β_4 1, sense- β_4 3: 5'-AACGATGCCACGCTGTTTC-3' and 5'-CGTGGGCGGAATACAAATGC-3', corresponding to nucleotides 76–94 and 118–137 of the identical strand of the β_4 gene sequence⁸, respectively. Each of the oligonucleotides can selectively hybridize with the complementary strand of the gene described; anti- β_2 2: 5'-TCGGCCCGCARGT-CRAAGAGG-3', corresponding to nucleotides 753–773 of the identical strand of the β_2 gene sequence⁶. It can selectively hybridize with the described β_2 mRNA; anti- β_3 3: 5'-CCTCCTCAGTTCAGATGTTT-3', corresponding to nucleotides 1,010–1,029 of the identical strand of the β_3 gene sequence⁷. It can selectively hybridize with the described β_3 mRNA; anti- β_4 2: 5'-ATATCCACAGCTTTGAGGAT-3', corresponding to nucleotides 618–638 of the identical strand of the β_4 gene sequence⁸. It can selectively hybridize with the described β_4 mRNA.

family of G_O composite subforms interacts with the corresponding receptor.

In addition to β_1 and β_3 subunits, β_2 - and β_4 -subtype subunits are expressed in GH₃ cells (Fig. 4). In contrast to the former subtypes, β_2 - and β_4 -subunit subtypes are not involved in the carbachol- or somatostatin-induced inhibition of voltage-dependent Ca²⁺ currents.

These data show that suppression of an individual β -subunit subtype blocks a particular signal transduction pathway between an agonist-activated receptor and the assigned effector. As the G protein in its heterotrimeric form binds to the activated receptor, the β -subunit suppression caused by injected antisense oligonucleotides prevented activation of a G protein and thus transduction of the signal across the membrane. We hypothesize that the receptor recognizes and binds individual G-protein subtypes, containing different β - and γ -subunits. Despite the likely existence of other G_O protein subtypes, the somatostatin and muscarinic (M₄) receptor functionally couple via G_{O2} containing the β_1 polypeptide and via G_{O1} containing β_3 , respectively, to the voltage-dependent Ca²⁺ channel. This hypothesis is consistent with data suggesting that the receptor recognition of G proteins is affected by specific $\beta\gamma$ complexes¹¹.

Although we cannot exclude a direct interaction between $\beta\gamma$ complexes and the Ca²⁺ channel, there are no indications for such effects. As β -subunit subtypes other than those targeted by the injected antisense oligonucleotide are apparently not affected, our results do not support the hypothesis of interchangeable β -subunits of $\beta\gamma$ complexes in intact cells¹, but rather suggest that the structural diversity of G proteins² corresponds to a functional one.

In summary, β -subunits appear to be an essential part of the specific G-protein coupling between a receptor and an effector.

On the basis of these and our other²¹ data, the subtypes of the G proteins involved in two signal transduction pathways in GH₃ cells can be specified as follows: muscarinic receptor → $\alpha_{O1}/\beta_3/\gamma$ → voltage-sensitive Ca²⁺ channel, and somatostatin receptor → $\alpha_{O2}/\beta_1/\gamma$ → voltage-sensitive Ca²⁺ channel. □

Received 19 December 1991; accepted 9 June 1992.

- Birnbaumer, L., Abramowitz, J. & Brown, A. M. *Biochim. biophys. Acta* **1031**, 163–224 (1990).
- Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802–808 (1991).
- Hekman, M. et al. *Eur. J. Biochem.* **169**, 431–439 (1987).
- Pinkas-Kramarski, R., Edelman, R. & Stein, R. *Neurosci. Lett.* **108**, 335–340 (1990).
- Codina, J., Stengel, D., Woo, S. L. C. & Birnbaumer, L. *FEBS Lett.* **207**, 187–192 (1986).
- Gao, B., Gilman, A. G. & Robishaw, J. D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6122–6125 (1987).
- Levine, M. A., Smallwood, P. M., Moen, P. T. Jr, Helman, L. J. & Ahn, T. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2329–2333 (1990).
- Von Weizsäcker, E., Strathmann, M. P. & Simon, M. I. *Biochem. biophys. res. Commun.* **183**, 350–356 (1992).
- Simonds, W. F., Butrynski, J. E., Gautam, N., Unson, C. G. & Spiegel, A. M. *J. biol. Chem.* **266**, 5363–5366 (1991).
- Cerione, R. A. et al. *Biochemistry* **26**, 1485–1491 (1987).
- Fawzi, A. B. et al. *J. biol. Chem.* **266**, 12194–12200 (1991).
- Kanaho, Y. et al. *J. biol. Chem.* **259**, 7378–7381 (1984).
- Tamir, H., Fawzi, A. B., Tamir, A., Evans, T. & Northup, J. K. *Biochemistry* **30**, 3929–3936 (1991).
- Fisher, K. J. & Aronson, N. N. Jr *Molec. cell. Biology* **12**, 1585–1591 (1992).
- Katada, T., Kusakabe, K., Oinuma, M. & Uli, M. *J. biol. Chem.* **262**, 11897–11900 (1987).
- Jelsema, C. L. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3623–3627 (1987).
- Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).
- Okabe, K. et al. *J. biol. Chem.* **265**, 12854–12858 (1990).
- Kim, D. et al. *Nature* **337**, 557–560 (1989).
- Boyer, J. L. *J. biol. Chem.* **264**, 13917–13922 (1989).
- Kleuss, C. et al. *Nature* **353**, 43–48 (1991).
- Law, S. F., Manning, D. & Reisine, T. *J. biol. Chem.* **266**, 17885–17897 (1991).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).

ACKNOWLEDGEMENTS. We thank M. I. Simon for providing the β_4 sequence before publication, K.-H. Jakobs for critically reading the manuscript, D. Wedlich for advice in immunochemical methods, and K. Spicher and K.-D. Hinsch for antibody generation and purification. Our work was supported by grants from Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie, and from Trüde-Görke-Stiftung.

Ectopic mesoderm formation in *Xenopus* embryos caused by widespread expression of a *Brachyury* homologue

V. Cunliffe & J. C. Smith

Laboratory of Developmental Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

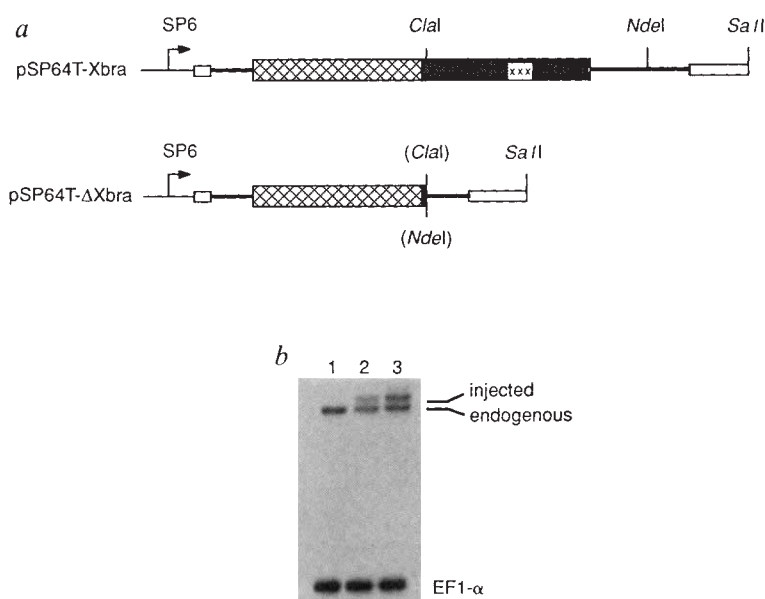
THE *Brachyury* (*T*) gene is required cell-autonomously for mesoderm formation in the posterior of the mouse embryo^{1,2}, and both its complementary DNA sequence and expression pattern³ closely resemble those of a *Xenopus* homologue (*Xbra*)⁴, suggesting that these genes have an evolutionarily conserved function in vertebrate development. Strong expression of *Xbra* messenger RNA is found in the ring of involuting mesoderm during *Xenopus* gastrulation⁴, and the expression of *Xbra* is an immediate-early response of animal pole blastomeres to mesoderm-inducing factors⁴. To assess the role of *Xbra* in mesoderm formation, we increased its domain of expression in the embryo by microinjection of *Xbra* transcripts into the animal pole of *Xenopus* embryos at the one-cell stage. We show that expression of *Xbra* by cells of the early embryo is sufficient to direct their development into differentiated mesodermal tissues. At the molecular level this response shows a sharp threshold of sensitivity to the dose of *Xbra* RNA delivered, and we suggest that *Xbra* may act as a genetic switch initiating posterior mesodermal specification during embryogenesis.

Full-length (*Xbra*) and mutant (Δ *Xbra*) *Xenopus Brachyury*

mRNA was prepared from the constructs shown in Fig. 1a and injected into the animal pole of *Xenopus* embryos at the one-cell stage. An RNase protection assay (Fig. 1b) revealed that the quantity of undegraded injected mRNA surviving to the gastrula stage was similar to the amount of endogenous *Xbra* mRNA. Both *Xbra*- and Δ *Xbra*-injected embryos began gastrulation normally, by forming dorsal lips and then completing blastopore formation. Δ *Xbra*-injected embryos closed their blastopores as mesoderm involution occurred and gastrulation was concluded, but *Xbra*-injected embryos failed to reduce the diameter of the blastopore, the first evidence of a disturbance in the normal process of gastrulation. Figure 2a shows that gastrulation perturbation occurs with a high frequency in *Xbra*-injected embryos and that it is specific, requiring the injection of full-length *Xbra* mRNA. Injection of similar amounts of *Xenopus* β -globin⁵, *Xenopus MyoD*⁶, or antisense-*Xbra* mRNAs also failed to affect embryogenesis noticeably. Histological sections of embryos at stage 13 showed that although all germ layers of the Δ *Xbra*-injected embryos were normal in structure and proportion, no involution of marginal zone mesoderm had occurred in *Xbra*-injected embryos, and a thick layer of tissue contiguous with and similar in appearance to marginal zone mesoderm had formed in the animal cap (Fig. 2b, c). This phenotype is very similar to that produced by intra-blastocoelic injection of the XTC mesoderm-inducing factor into late blastulae, where a continuous layer of mesoderm is formed in the animal hemisphere⁷. *Xbra*-injected embryos collected at stage 12 expressed elevated levels of muscle-specific actin⁸ mRNA in comparison to the low quantity in uninjected and Δ *Xbra*-injected embryos at this stage (Fig. 2d), suggesting that there was an increase in the proportion of mesoderm being formed.

The mesodermal character of this animal cap tissue in *Xbra*-injected embryos was firmly established by isolating animal caps

FIG. 1 Construction of plasmids for synthesis of *Xbra* and mutant *Xbra* mRNAs and measurement of the abundance of injected and endogenous *Xbra* mRNAs in the mid-gastrula. a, A full-length *Xbra* cDNA clone⁴ was inserted into the expression vector pSP64T (ref. 5) between the 5' and 3' untranslated sequences of the *Xenopus* β -globin gene to yield pSP64T-*Xbra*. A mutant *Xbra* cDNA (Δ *Xbra*) was derived from this plasmid by removal of *Xbra* sequences between the unique *Clal* and *NdeI* sites. The DNA was polished with Klenow fragment and religated to yield pSP64T- Δ *Xbra*. Narrow lines, vector sequences; small rectangles, β globin 5' and 3' untranslated regions; bold lines, *Xbra* 5' and 3' untranslated regions; large rectangle, *Xbra* translated region; hatched area, N-terminal region displaying >85% identity with mouse *Brachyury* gene; shaded area, less well conserved region containing conserved potential N-linked glycosylation sites (denoted by XXX). For *in vitro* transcription, plasmids were linearized with *SaI* and transcribed with SP6 polymerase. *Xbra* and Δ *Xbra* mRNAs produced in this way are translated in rabbit reticulocyte lysate and produce proteins of expected molecular masses (48K and 26K, respectively, data not shown). b, A ribonuclease (RNase) protection assay shows that after injection of 5 ng synthetic RNA at the one-cell stage, the amount of injected *Xbra* or Δ *Xbra* RNA remaining undegraded in the mid-gastrula is similar to the level of the endogenous *Xbra* mRNA, and thus in the physiological range. An accurate assessment of the dose of *Xbra* protein responsible for the effects described in this paper will require *Xbra*-specific antibodies. Lane 1, uninjected embryos; lane 2, *Xbra*-injected embryos; lane 3, Δ *Xbra*-injected embryos. For each sample, RNA was isolated from 20 stage-11.5 embryos and 6 μ g analysed for expression of *Xbra* and EF1 α transcripts as described in ref. 4. Incorporation of a short plasmid polylinker sequence into the synthetic transcripts and the *Xbra* probe created a size difference between the probe lengths protected by the injected and endogenous *Xbra* mRNAs, allowing quantitation of relative abundance. *In situ* hybridizations to individual embryos (data not shown) indicated that the injected mRNA was distributed uniformly around the blastocoel roof, and that there was some variation between embryos in the relative abundance of injected and endogenous *Xbra* mRNAs.



METHODS. Standard recombinant DNA manipulation techniques were used for template construction and preparation²⁰, and *in vitro* transcriptions were according to manufacturer's recommendations (Promega). Embryo preparation and media (NAM, Normal Amphibian Medium) were as in ref. 21. During injection, embryos were cultured in 75% NAM, 5% Ficoll 400 (Pharmacia). In all experiments (unless stated) 5 ng synthetic RNA was injected in a volume of ~10 nl into the animal pole of *Xenopus* embryos at the one-cell stage using a Singer microinjection apparatus. This quantity is similar to the amounts of synthetic mRNA used for microinjection in previously published experiments^{6,17,18,22}. When embryos reached stage 6 (ref. 23) the medium was gradually replaced with 10% NAM before culture overnight or with 75% NAM before animal cap dissection.

and performing on these explants a combination of RNase protection assays for mesoderm and ectoderm markers, histological analysis, and by immunostaining with a muscle-specific antibody. *Xsna* is a zinc-finger gene whose expression is induced in *Xenopus* mesoderm and in animal caps treated with mesoderm-inducing factors⁹. Figure 3a shows that animal caps dissected from *Xbra*-injected embryos at stage 8 expressed elevated levels of *Xsna* transcripts when cultured to stage 12, in comparison to the low level observed in caps from uninjected and $\Delta Xbra$ -injected embryos. Moreover, explants from *Xbra*-injected embryos cultured to stage 23 showed both strong expression of muscle-specific actin mRNA⁸ and a greatly reduced level of expression of epidermal keratin mRNA¹⁰, indicating a switch from ectodermal to mesodermal fate (Fig. 3a). Histological sections of *Xbra*-injected caps collected at stage 37 revealed the presence of muscle masses, mesothelium and mesenchyme (Fig. 3b). These tissue types are often seen together in caps treated with high doses of basic fibroblast growth factor (bFGF) or low doses of activin¹¹. Immunostaining with the muscle-specific monoclonal antibody 12/101 (ref. 12) confirmed the presence of differentiated muscle tissue (Fig. 3c).

T/T homozygous mouse embryos fail to generate mesoderm posterior to somite 7 but do develop mesodermal tissues anterior to this boundary². This suggests that the role of *Xbra* in mesodermal specification might be restricted along the antero-posterior axis. Figure 3d shows that *Xbra* mRNA injection caused marked induction of *Xhox3*, a homeobox-containing gene that is strongly activated by bFGF and most highly expressed in posterior mesoderm¹³, but no induction of *goosecoid*, a homeobox-containing gene induced by activin and expressed in notochord and anterior mesoderm¹⁴. These results imply that *Xbra* causes posterior mesoderm formation and that it acts upstream of *Xhox3*.

The types of mesoderm generated by activin and by bFGF are concentration-dependent¹¹, and a previous study of the morphogenetic properties of activin showed that the threshold for induction of the muscle-specific actin gene in single cells from animal cap explants was very sharp, occurring over a 1.5-fold change in activin concentration¹⁵. A similarly sharp threshold effect was observed when a twofold dilution series of *Xbra* mRNA was injected into embryos and muscle-specific actin expression analysed at stage 20 (Fig. 3e). Although a concentration of 0.13 ng nl⁻¹ failed to induce, strong expression was elicited on injection of RNA at 0.25 ng nl⁻¹, and the level of induction remained constant when higher concentrations of 0.5 and 1.0 ng nl⁻¹ were injected. Although injection of *XMyoD* RNA neither disrupts development nor causes myogenesis, it does activate muscle-specific actin genes⁶, indicating that it may act downstream of *Xbra*. Moreover, increasing doses of *XMyoD* mRNA generated a more gradual increase in the level of muscle-specific actin mRNA in isolated animal caps⁶, in contrast to the sharp threshold of sensitivity of the muscle-specific actin gene to increasing concentrations of *Xbra* mRNA. This again suggests that these two molecules play distinct roles in the processes leading to muscle gene activation.

These results show that inappropriate expression of *Xbra* mRNA in the animal cap of *Xenopus* embryos causes the formation of ectopic mesoderm, as predicted from the histology of *Xbra*-injected embryos, and that this mesoderm has posterior character. It is possible, therefore, that *Xbra* mediates the effects of low concentrations of activin, high concentrations of bFGF, or injection of bone morphogenetic protein-4 mRNA¹⁶, all of which treatments induce mesoderm of ventro-posterior character. The mesoderm-inducing effect of *Xbra* differs from the actions of other cell-autonomously acting *Xenopus* genes, such as *Xhox3* (ref. 17) and *XIHbox6* (ref. 18), where injection of

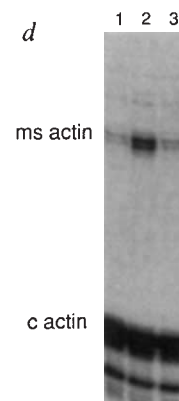
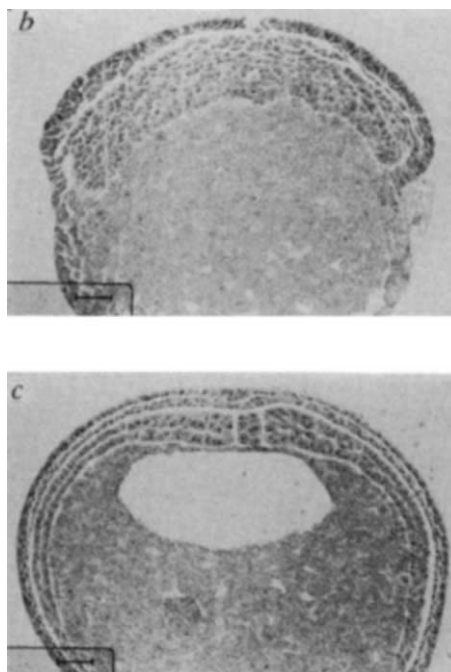
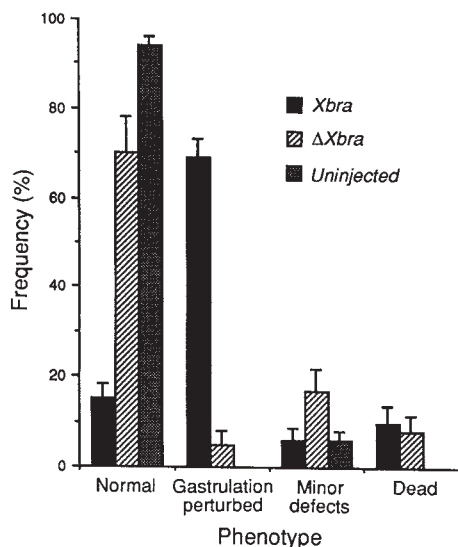


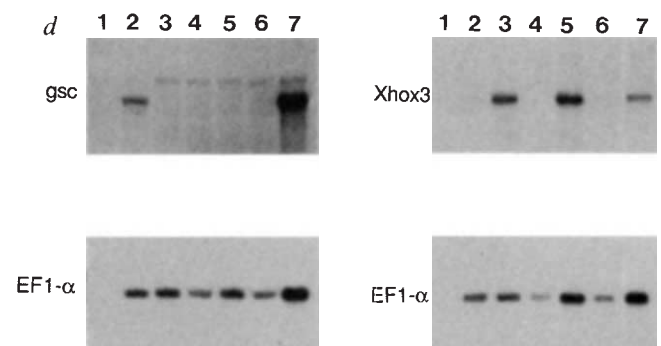
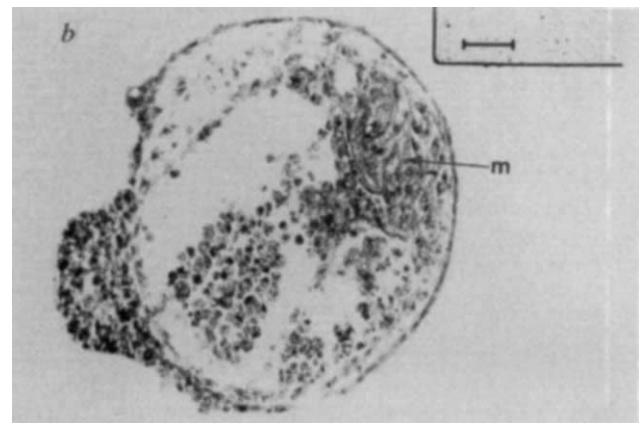
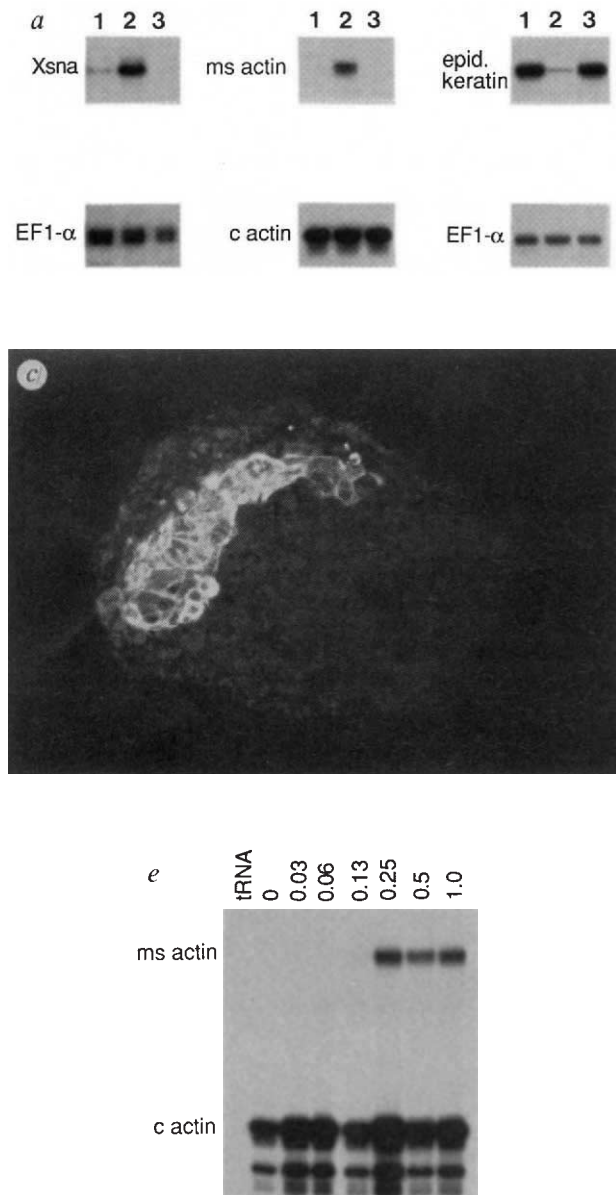
FIG. 2 Effect of injection of *Xbra* and $\Delta Xbra$ mRNAs on development of *Xenopus* embryos. *a*, Embryos were scored when uninjected controls reached tailbud stage and each classified into one of four categories: normal; gastrulation perturbed (no mesoderm involution at blastopore, retracted and wrinkled deeply pigmented outer layer); minor defects (small head, cyclopia, incomplete closure of neural tube posteriorly); and dead. Histogram represents aggregate results of nine separate experiments. Total numbers of embryos injected: *Xbra*, 396; $\Delta Xbra$, 396; uninjected controls, 451. Values are means and error bars represent s.e.m. *b*, *c*, Histology at control stage 13 of embryos injected with *Xbra* (vertical section) (*b*), $\Delta Xbra$ RNA (transverse section) (*c*).

A thick pad of tissue histologically similar to and contiguous with marginal zone mesoderm is clearly visible in the animal cap of *b*. Scale bars, 100 μ m. *d*, An RNase protection assay shows that at control stage 12 *Xbra*-injected embryos express elevated levels of muscle-specific actin (ms actin) as compared with uninjected and $\Delta Xbra$ -injected embryos. Cytoplasmic actin (c actin) is an effective internal control. Lane 1, $\Delta Xbra$ -injected embryos; lane 2, *Xbra*-injected embryos; lane 3, uninjected embryos. METHODS. Specimens were fixed in formalin and embedded in paraffin wax¹¹. Sections (7 μ m) were cut and stained with a combination of Feulgen, light green and orange G as in ref. 24. RNase protections were done as in ref. 15.

RNA perturbs development but has not been reported to induce mesoderm from animal pole tissue. Thus, where the phenotype of *T/T* mouse embryos indicates that *Brachyury* is necessary for posterior mesoderm formation, our experiments in *Xenopus* show that it is sufficient for generating these cell types. The absence of dorso-anterior mesoderm in response to *Xbra*, despite the expression of *Xbra* in the precursors of these tissues, suggests that the specification of such cells may require the activity of other genes which are, perhaps, induced by high concentrations of factors such as activin. These genes may

modulate some of the effects of *Xbra*, for although activin and bFGF both induce expression of *Xbra* to a similar level⁴, activin is a much stronger inducer of muscle-specific actin than bFGF¹¹. One way to investigate this question will be to coinject *Xbra* RNA with RNA of other immediate-early response genes.

Finally, the existence of a sharp threshold for muscle gene activation in response to different doses of *Xbra* RNA implies, along with the loss of function phenotype of *T/+* mice, that cells are very sensitive to the precise concentration of *Xbra* protein they experience. In this respect the activation of actin



mesothelium, which are regarded as ventral and lateral mesodermal cell types¹¹. Scale bar, 25 μ m. Caps from Δ *Xbra*-injected embryos form atypical ciliated epidermis identical to those from control uninjected embryos (data not shown). c, Immunostaining with muscle-specific monoclonal antibody 12/101 of animal cap section from *Xbra*-injected embryo, collected at control stage 37. d, RNase protection assays show that animal caps from *Xbra*-injected embryos have properties of posterior but not anterior mesoderm. Lanes 1, transfer RNA control; 2, uninjected caps incubated with 50 U ml⁻¹ activin A; 3, uninjected caps incubated with 50 U ml⁻¹ bFGF; 4, uninjected caps; 5, *Xbra*-injected caps (2.5 ng per embryo); 6, Δ *Xbra*-injected caps (2.5 ng per embryo); 7, stage 12.5 whole embryo (see ref. 25 for definition of units of activin and bFGF activity). Animal caps were cut from embryos at stage 8 and incubated until control stage 12.5 in the presence or absence of peptide growth factor before collecting and analysing for *gooseoid* (*gsc*) mRNA (left panel) and *Xhox3* mRNA (right panel). In both panels the loading control was EF1- α . e, Dose-response profile of animal cap explants from embryos injected with a twofold dilution series of *Xbra* RNA. Lanes 1, tRNA control; 2, uninjected; 3, 0.03 ng ml⁻¹ *Xbra* RNA; 4, 0.06 ng ml⁻¹ *Xbra* RNA; 5, 0.13 ng ml⁻¹ *Xbra* RNA; 6, 0.25 ng ml⁻¹ *Xbra* RNA; 7, 0.5 ng ml⁻¹ *Xbra* RNA; 8, 1.0 ng ml⁻¹ *Xbra* RNA.

METHODS. RNAs were injected in a volume of ~10 nl as described in the legend to Fig. 1. Animal caps were dissected from embryos at uninjected control stage 8 in 75% NAM and cultured to the appropriate stage. Where peptide growth factors were used 0.1% BSA was added to 75% NAM. RNA was isolated as described²⁰, hybridized with radioactive RNA probes detecting *Xsna* (ref. 9), muscle-specific actin⁸, epidermal keratin¹⁰, *gooseoid* (ref. 14) or *Xhox3* (ref. 13), and hybrids analysed as in ref. 15. Histological analysis of animal caps was done as for whole embryos. Immunohistochemistry was done on trichloroacetic acid-fixed frozen sections of animal caps with monoclonal antibody 12/101 as described¹⁵.

FIG. 3 Formation of ectopic mesoderm in animal caps isolated from *Xbra*-injected embryos. a, RNase protection assays show that a programme of mesodermal gene expression is initiated by *Xbra*-injected prospective ectoderm and that ectodermal differentiation is suppressed. Lanes 1, animal caps from uninjected embryos; lanes 2, animal caps from *Xbra*-injected embryos; for *Xsna* analysis, caps were collected at control stage 12 and an EF1- α probe⁴ was used as loading control. For muscle-specific actin (ms actin), caps were collected at control stage 23; cytoskeletal actin (c actin) is the loading control. For epidermal keratin (epid. keratin), caps were collected at control stage 23 and the EF1- α probe was the loading control. b, Histological sectioning of animal caps injected with *Xbra* RNA, collected at control stage 37. In caps from *Xbra*-injected embryos, a muscle mass (m) is indicated by the arrow, and is surrounded by concentric rings of mesenchyme and

genes by *Xbra* resembles the cooperative activation of *hunchback* by *bicoid* in *Drosophila*¹⁹, and it will be of great interest to establish whether *Xbra* interacts directly with regulatory elements of mesoderm-specific genes. □

Received 12 February; accepted 15 June 1992.

- Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. *Nature* **343**, 617–622 (1990).
- Rashbass, P., Cooke, L. A., Herrmann, B. G. & Beddington, R. *Nature* **353**, 348–351 (1991).
- Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. *Nature* **343**, 657–659 (1990).
- Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. *Cell* **67**, 79–87 (1991).
- Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057–7071 (1984).
- Hopwood, N. D. & Gurdon, J. B. *Nature* **347**, 197–200 (1990).
- Cooke, J. & Smith, J. C. *Dev. Biol.* **131**, 383–400 (1989).
- Mohun, T. J., Brennan, S., Dathan, N., Fairman, S. & Gurdon, J. B. *Nature* **311**, 716–721 (1984).
- Sargent, M. G. & Bennett, M. F. *Development* **109**, 967–973 (1990).
- Jonas, E., Sargent, T. D. & Dawid, I. B. *Proc. natn. Sci. U.S.A.* **82**, 5413–5417 (1985).
- Green, J. B. A., Howes, G., Symes, K., Cooke, J. & Smith, J. C. *Development* **108**, 173–183 (1990).

- Kinter, C. R. & Brookes, J. P. *Nature* **308**, 67–69 (1984).
- Ruiz i Altaba, A. & Melton, D. A. *Development* **106**, 173–183 (1989).
- Cho, K. W. Y., Blumberg, B., Steinbesser, H. & De Robertis, E. M. *Cell* **67**, 1111–1120 (1991).
- Green, J. B. A. & Smith, J. C. *Nature* **347**, 391–394 (1990).
- Dale, L., Howes, G., Price, B. M. J. & Smith, J. C. *Development* **115**, 573–585 (1992).
- Ruiz i Altaba, A. & Melton, D. A. *Cell* **57**, 317–326 (1989).
- Cho, K. W. Y., Morita, E. A., Wright, C. V. E. & De Robertis, E. M. *Cell* **65**, 55–64 (1991).
- Driever, W. & Nusslein-Volhard, C. *Nature* **337**, 138–143 (1989).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1982).
- Slack, J. M. W. *J. Embryol. exp. Morph.* **80**, 289–319 (1984).
- Amaya, E., Musci, T. J. & Kirschner, M. W. *Cell* **66**, 257–270 (1991).
- Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis (Daudin)* (North-Holland, Amsterdam, 1967).
- Cooke, J. *J. Embryol. exp. Morph.* **51**, 165–182 (1979).
- Cooke, J., Smith, J. C., Smith, E. J. & Yaqoob, M. *Development* **101**, 893–908 (1987).

ACKNOWLEDGEMENTS. We thank B. Price for subcloning, W. Hatton for histology, M. Saha for the *Xhox3* plasmid, E. De Robertis for the *gsc* plasmid, and the NIMR Photographic Department for photographic work. V.C. is the recipient of a postdoctoral fellowship from the SERC.

An unusual feature revealed by the crystal structure at 2.2 Å resolution of human transforming growth factor- β 2

Michael P. Schlunegger & Markus G. Grütter

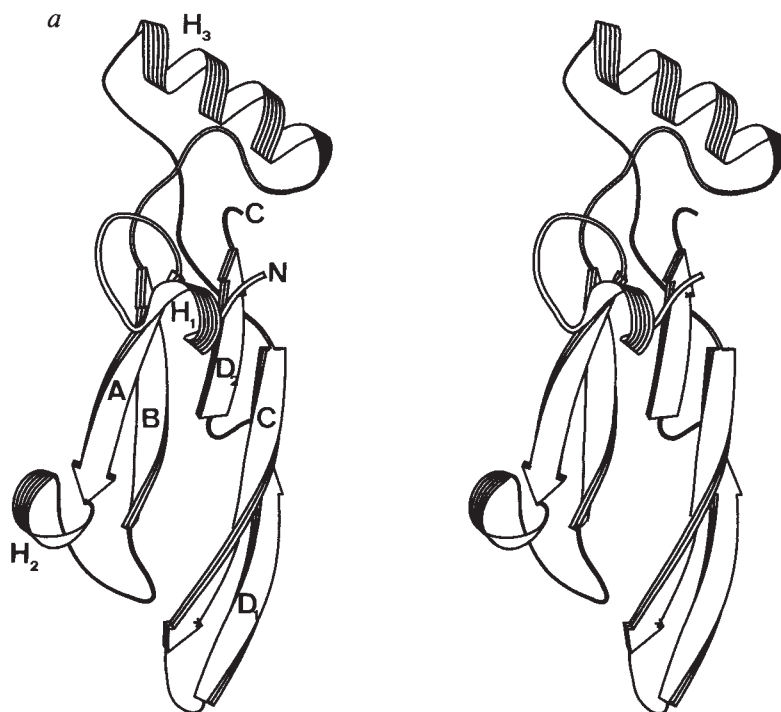
Department of Biotechnology, Pharmaceuticals Division, Ciba-Geigy, CH-4002 Basel, Switzerland

TRANSFORMING growth factor type β 2 (TGF- β 2)¹ is a member of an expanding family of growth factors that regulate proliferation and differentiation of many different cell types^{2,3}. TGF- β 2 binds to various receptors⁴, one of which was shown to be a serine/threonine kinase⁵. TGF- β 2 is involved in wound healing⁶, bone formation⁷ and modulation of immune functions⁸. We report here the crystal structure of TGF- β 2 at 2.2 Å resolution, which reveals a novel monomer fold and dimer association. The monomer consists of two antiparallel pairs of β -strands forming a flat curved surface and a separate, long α -helix. The disulphide-rich core has one disulphide bond pointing through a ring formed by the sequence motifs Cys-Ala-Gly-Ala-Cys and Cys-Lys-Cys, which are themselves connected through the cysteines. Two monomers are connected through a single disulphide bridge and associate such that the helix of one subunit interacts with the concave β -sheet surface of the other. Four exposed loop regions might determine receptor specificity. The structure provides a suitable model for the TGF- β s and other members of the superfamily^{9–11} and is the basis for the analysis of the TGF- β 2 interactions with the receptor.

Human recombinant TGF- β 2 was prepared and crystallized as described earlier¹². Table 1 describes data collection, structure solution and refinement of the TGF- β 2 structure. The TGF- β 2 monomer is a flat, elongated and slightly bent molecule with an overall size of $60 \times 20 \times 15$ Å³. The unusual monomer fold consists at one end of a long α -helix with its axis perpendicular to the β -sheet and which is separated from two antiparallel pairs of β -strands (Fig. 1a). The core of the molecule has an additional unusual fold involving the four intramolecular disulphide bridges (Fig. 1b). The amino-terminal short α -helix on the outside of the monomer is followed by an exposed loop that is fixed to the core through a disulphide bridge. The other three disulphide bonds connect the different β -strands with each other. Two of these disulphide bridges form a narrow and rigid eight-membered ring consisting of the residues Cys 44-Ala 45-Gly 46-

Ala 47-Cys 48-Cys 111-Lys 110-Cys 109-(Cys 44). The disulphide bridge, Cys 15 to Cys 78, connecting the N terminus with the β -strand marked C in Fig. 1a, points directly through this ring. We will refer to this unusual structural feature as the TGF- β 'knot'. All the intramolecular disulphide bridges are forming a tight and compact core. Similar eight-membered rings formed by two disulphide-connected backbones have also been found in endothelins, sarafatoxins, bee-venom toxins and scorpion-venom toxins¹³. In these proteins the Cys-X-X-X-Cys sequence motif is part of an α -helix and the disulphide bonds are used to stabilize this helix. A similar but larger knot to that in TGF- β 2 was found in the potato inhibitor of carboxypeptidase A (ref. 14). Its sequence motifs Cys-X-X-X-Cys and Cys-X-X-Cys form a nine-membered ring through which a third disulphide bridge is pointing. Comparison of the van der Waals surfaces of the two knots shows that the eight-membered ring of the TGF- β knot is the most compact ring through which a disulphide bond can pass.

The monomer has a solvent-accessible surface area of 4,400 Å² (ref. 15). In the dimer (Fig. 1c), two monomers are connected by a disulphide bridge (Cys 77) lying exactly on the crystallographic 2-fold axis ($x-y$, $-y$, $1/3-z$). The surface area of the dimer is 7,000 Å², indicating that about 1,800 Å² (900 Å² per

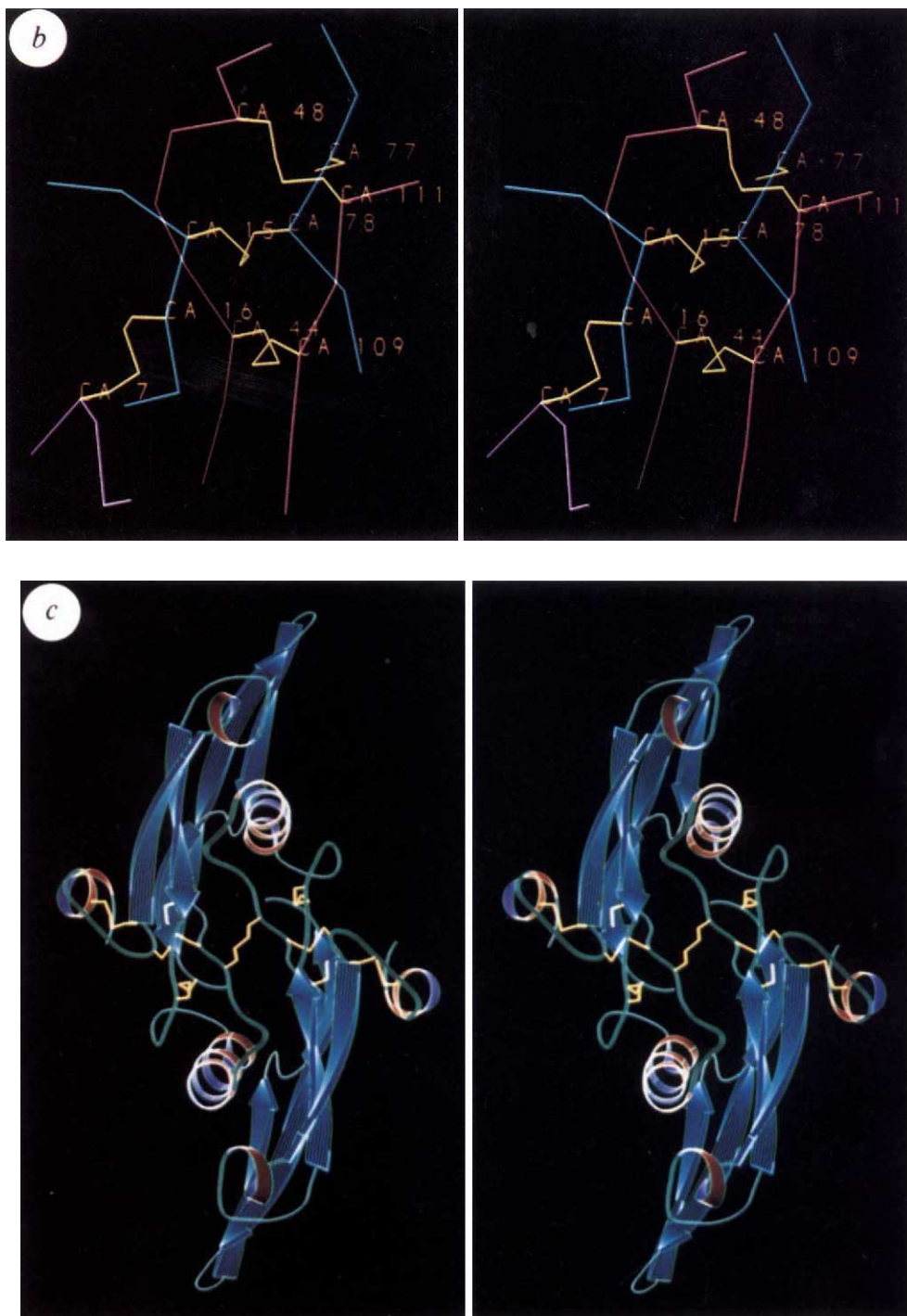


monomer) are buried in the dimer. The long axes of the monomer molecules are roughly perpendicular to the 2-fold axis. The long α -helices H_3 are in tight contact with the β -sheet of the other subunit. Most of the interactions between residues 56 to 65 of the α -helix H_3 of one subunit and the β -sheet of the second subunit are made between hydrophobic and aromatic residues that are tightly packed and buried in the dimer interface. Additional stabilization of this helix-sheet interaction is provided by two hydrogen bonds between the two subunits: the carbonyl oxygens of Asn 42 and Asn 103 of one subunit to the side-chain of His 58 and to the main-chain nitrogen of His 58 of the other subunit, respectively. The packing of the two monomers is less compact in the region of the intermolecular disulphide bond.

Hydrophobic interactions between Cys 77 of one subunit and Val 79 of the second subunit, as well as an extended hydrogen-bonding network in which several clearly defined water molecules are involved, also contribute to the stabilization of the dimer. Predominantly hydrophobic interactions between the monomers are found between Phe 43 (β -strand B; Fig. 1g) and Ala 72, Ser 73 and Ala 74 (β -strand C of the other subunit).

Possible receptor-binding sites include most of the likely side-chains of surface residues located in protruding loops. Sequence variability in TGF- β 1 to β 5 in such regions must be responsible for the specificity for the different receptor types. Viewed along the 2-fold axis, the biologically active dimer of TGF- β 2 (Fig. 1c) shows three or four such regions (Fig. 2a, b).

FIG. 1 a (opposite), Stereo RIBBON²⁴ picture of the TGF- β 2 monomer fold. An N-terminal short α -helix H_1 (residues 3–8) on the outside of the monomer is followed by an exposed loop, which is fixed to β -strand A (residues 16–23) through a disulphide bridge. The second β -strand B (residues 38–45) is connected by a long loop of 14 amino acids, including a small helical turn H_2 (residues 24–28), to β -strand A. The β -strand B is fixed to the C terminus by two disulphide bonds. The α -helix H_3 (residues 56–69) of nearly four turns, situated at one end of the molecule, connects β -strand B and the long β -strand C (residues 82–91). β -strands C and D_1 (residues 95–102) are connected by a β -hairpin. Pro 85, in β -strand C, disrupts the hydrogen-bonding pattern between β -strands C and D_1 and forces β -strand D_1 to wind around β -strand C in order to continue hydrogen-bonding with the opposite side of β -strand C (β -strand D_2 , residues 105–110). b, Stereo picture showing parts of the core displaying all the cysteines and parts of the backbone. The two disulphide bridges Cys 44–Cys 109 and Cys 48–Cys 111 (backbone coloured in red) are forming a tight ring of eight residues. The disulphide bridge Cys 15–Cys 78 (backbone coloured in blue) is pointing directly through this ring. The N terminus (backbone coloured in purple) is fixed by the disulphide bridge Cys 7–Cys 16. Cys 77 (right upper corner of the picture) forms the only intermolecular disulphide bond. The rigidity is emphasized by low individual atomic temperature factors in the range of 5 to 25 Å². c, Stereo RIBBON²⁴ picture of the TGF- β 2 dimer viewed along the crystallographic 2-fold axis. The centre of the single intermolecular disulphide bond between Cys 77 of the first subunit and Cys 77* of the second subunit is lying exactly on this 2-fold axis. All the intra- as well as the intermolecular disulphide bridges are in yellow.



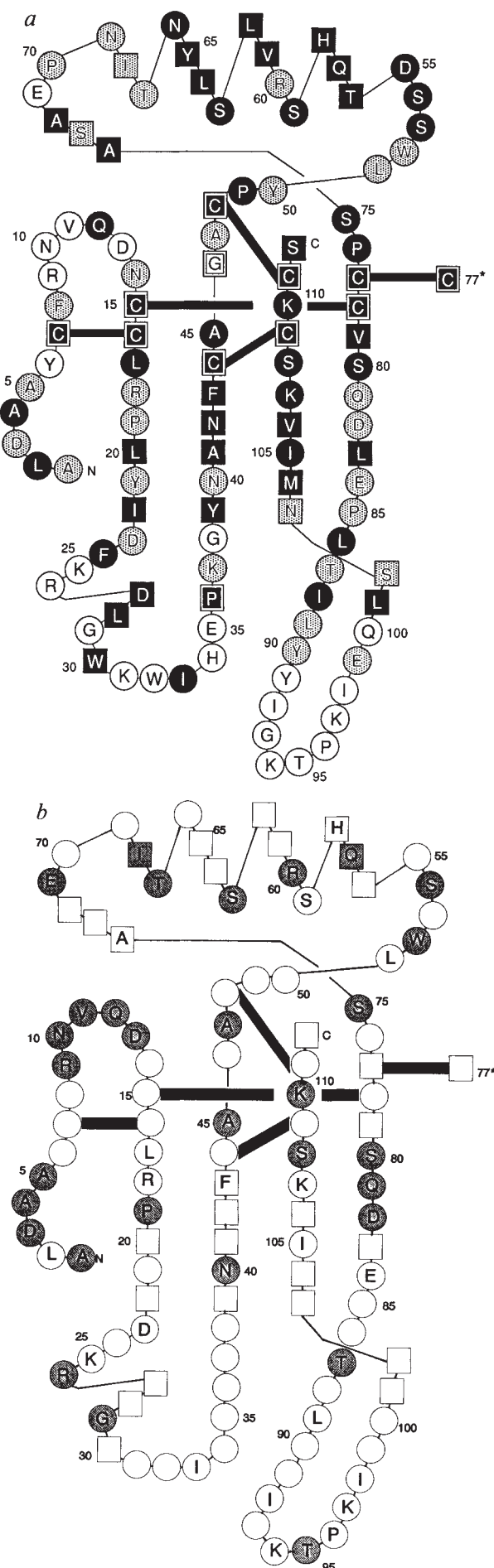


FIG. 2 a, Schematic representation of the TGF- β 2 monomer highlighting structurally important residues. Amino acids are indicated by their one-letter code. Double squares, invariant residues for all TGFs and related proteins (Cys 7 and Cys 16 are only invariant in the five TGF- β isoforms). Characterization by the side-chain solvent-accessible area A^{25} : not shaded, exposed residues ($A > 60 \text{ \AA}^2$); light shaded, intermediately exposed residues ($20 \text{ \AA}^2 < A < 60 \text{ \AA}^2$); dark shaded, buried residues ($A < 20 \text{ \AA}^2$). The single squares represent residues involved in the dimer interface having a distance of less than $4 \text{ \AA}$ from the other subunit. Glycine residues have not been included in the solvent accessibility analysis. **b**, Schematic representation of the TGF- β 2 monomer highlighting residues that show changes in the five isoforms TGF- β 1 to TGF- β 5. Squares, residues involved in the dimer interface (as in **a**). Not shaded, residues that show conservative changes; shaded, residues that show strong changes. Residues not indicated with their one-letter code are invariant in all five isoforms. The insertion of two additional amino acids in TGF- β 4 after Val 11 is not shown. **c**, Sequence alignment of the mature or proposed mature region of 20 TGF- β -like proteins beginning with the first invariant cysteine residue (which is Cys 15 and not Cys 16 of TGF- β 2 as was indicated earlier²⁵). The conserved residues, Cys 16, Pro 36, Cys 44, Gly 46, Cys 48, Cys 77, Cys 78, Cys 109 and Cys 111 are indicated by boxes. The superfamily has been separated into four subgroups based on the relative sequence similarities between these members²⁶. TGF- β subgroup consists of human TGF- β 1, TGF- β 2 and TGF- β 3, chicken TGF- β 4 and *Xenopus* TGF- β 5. Vg/dpp subgroup consists of the decapentaplegic gene product dpp of *Drosophila*, Vg-1 of the vegetal pole of *Xenopus*, mouse Vgr-1, the mammalian Vg-1 homologue, mouse growth/differentiation factor 1 (GDF-1), *Drosophila* 60A, human bone morphogenetic proteins BMP-2a, BMP-2b/BMP-4, BMP-3, BMP-5, BMP-6 and BMP-7/OP-1 (osteogenic protein 1). The inhibin subgroup consists of porcine inhibin α , β A and β B. The MIS subgroup contains a single member, human Müllerian inhibitory substance MIS (refs 25, 26).

C TGF- β subgroup

TGF- β 1	C	CVRQ	LYIDFRKDLGWK	WIHE	P	KGYNANF	C	L	G	P	C	P	YIWI	---	SLD	---	TQYSK
TGF- β 2	C	CLRP	LYIDFKRDLGWK	WIHE	P	KGYNANF	C	A	G	A	C	P	YIWI	---	SSD	---	TQHSR
TGF- β 3	C	CVRP	LYIDFRKDLGWK	WVHE	P	KGYNANF	C	S	G	P	C	P	YIWI	---	SAD	---	TTTST
TGF- β 4	C	CVRP	LYIDFRKDLGWK	WIHE	P	KGYNANF	C	M	G	P	C	P	YIWI	---	SAD	---	TQYSK
TGF- β 5	C	CVKP	LYINFRKDLGWK	WIHE	P	KGYEANY	C	L	G	N	C	P	YIWI	---	SMD	---	TQYSK
		15			36			44	46	48							

Vg/dpp subgroup

dpp	C	-RRHSLYVDF-SDVGWDDWIVA	P	LGVDAYY	C	H	G	K	C	P	-FPLADHFN	---	TNHAI
Vg-1	C	-KKRHLVYEF-KDVGWQNVIVA	P	QGYMANY	C	Y	G	E	C	P	-YPLTEILNG	---	SNHAI
Vgr-1	C	-KKHSLVVSF-QDLGWQDWIIA	P	KGYAANY	C	D	G	E	C	P	-SFPLNAMHNA	---	TNHAI
GDF-1	C	-RRRLHVSF-REVGWHRWIVA	P	RGLFANP	C	Q	G	T	C	ALPETLRGPGGPPALNHAI			
60A	C	-QMOTLYIDF-KDLGWHDWIIA	P	EGYGAFY	C	S	G	E	C	N	-FPLNAMHNA	---	TNHAI
BMP-2a	C	-KKRHLVYDF-SDVGWDDWIVA	P	PGYHAFY	C	H	G	E	C	P	-FPLADHFN	---	TNHAI
BMP-2b/BMP-4	C	-RRHSLVYDF-SDVGWDDWIVA	P	PGYQAFY	C	H	G	D	C	P	-FPLADHFN	---	TNHAI
BMP-3	C	-ARRYLKVSF-ADIGWSEWIIA	P	KSFDAFY	C	S	G	A	C	QFPMPSKSLKPS	---	NHAI	
BMP-5	C	-KKHSLVVSF-RDLGWQDWIIA	P	EGYAIFY	C	D	G	E	C	S	-FPLNAMHNA	---	TNHAI
BMP-6	C	-KKHSLVVSF-QDLGWQDWIIA	P	KGYAANY	C	D	G	E	C	S	-FPLNAMHNA	---	TNHAI
BMP-7/OP-1	C	-KKHSLVVSF-RDLGWQDWIIA	P	EGYAIFY	C	E	G	E	C	A	-FPLNSYMNA	---	TNHAI

Inhibin subgroup

Inhibin α	C	-HRVALNISF-QELGWERWIVY	P	PSFIFHY	C	H	G	C	GLHIPPNLSLPV--PGAPP	
Inhibin β A	C	--KKQFFVSF-KDIGWNDWIIA	P	SGYHANY	C	E	G	E	C	PSHIAGTSGSSL--SFHST
Inhibin β B	C	--RQFFIDF-RLIGWNDWIIA	P	TGYGNY	C	E	G	S	C	PAYLAGVPGSAS--SFHTA

MIS subgroup

MIS	C	ALRELSVDLRAE	---	RSVLI	P	ETYQANN	C	Q	G	V	C	GWPQSDRNPY	---	GNH-V
-----	---	--------------	-----	-------	---	---------	---	---	---	---	---	------------	-----	-------

TGF- β subgroup

TGF- β 1	V	LALYN	---	QHNPGASAP	CC	V	---	PQALEPLPIVY	---	VGRKPKV	---	EQLSNMIVRS	C	K	C	S
TGF- β 2	V	LALYN	---	TINPEASAP	CC	V	---	SQDLEPLTILY	---	IGKTPKI	---	EQLSNMIVRS	C	K	C	S
TGF- β 3	V	LGLYN	---	TLNPEASAP	CC	V	---	PQDLEPLTILY	---	VGRTPKV	---	EQLSNMIVRS	C	K	C	S
TGF- β 4	V	LALYN	---	QHNPGASAP	CC	V	---	PQDLEPLTILY	---	VGRNVRV	---	EQLSNMIVRS	C	K	C	S
TGF- β 5	V	LALYN	---	QHNPGASAP	CC	V	---	PQDLEPLTILY	---	VGRNVRV	---	EQLSNMIVRS	C	K	C	S
								77/78					109	111		

Vg/dpp subgroup

dpp	V	QTLVNN	---	MNPGKVPKA	CC	V	---	PTQLDSVAMLYLND	---	QSTVVLKNYQEMTVVG	C	G	C	R
Vg-1	V	LQTLVNS	---	IEPEDILP	CC	V	---	PTKMSPIMLFY	---	DNNDNVLLRHYENNAVDE	C	G	C	R
Vgr-1	V	QTLVHL	---	MNPEYVKKP	CC	A	---	PTKLNIAISVLYF	---	DDNSNVILKKYRNMMVRA	C	G	C	H
GDF-1	V	LRALMHAAA	---	PTPGAGSP	CC	V	---	PERLSPIVSLFF	---	DNNDNVLLRHYEDMNVDE	C	G	C	R
60A	V	QTLVHL	---	LEPKKVPKP	CC	A	---	PTRLGALPVLYL	---	LDNENNVILKKYRNMMVRA	C	G	C	H
BMP-2a	V	QTLVNS	---	VNSKIPKA	CC	V	---	PTLSAISMLYL	---	DENKVVLLKNYQDMVVEG	C	G	C	R
BMP-2b/BMP-4	V	QTLVNS	---	VNSKIPKA	CC	V	---	PTLSAISMLYL	---	DEYDKVLLKNYQEMTVVG	C	G	C	R
BMP-3	V	IQSIVRA	---	VGVVPGIPEP	CC	V	---	PEKMSLSLILFF	---	DENKVVLLKNYQDMVVEG	C	A	C	R
BMP-5	V	QTLVHL	---	MFPDHVPKP	CC	A	---	PTKLNIAISVLYF	---	DDNSNVILKKYRNMMVRA	C	G	C	H
BMP-6	V	QTLVHL	---	MNPEYVKKP	CC	A	---	PTKLNIAISVLYF	---	DDNSNVILKKYRNMMVRA	C	G	C	H
BMP-7/OP-1	V	QTLVHF	---	INPETVKKP	CC	A	---	PTQLNIAISVLYF	---	DDNSNVILKKYRNMMVRA	C	G	C	H

Inhibin subgroup

Inhibin α	T---	PAQPYSLPGAQP	CC	AALPGTMRPLHVRTSDGGYSFKYETVPLNLTQH	C	A	C	I
Inhibin β A	VINHRYMRGHSPFANLKS	CC	V--PTKLRPMSMLYY--DDGQNIKKDIQNMIVEE	C	G	C	S	
Inhibin β B	VVNQYRMRLNPGT-VNS	CC	I--PTKLSTMSMLYF--DDEYNIVKRDVPMIVEE	C	G	C	A	

MIS subgroup

MIS	V	LLLL	---	KMQARGAALARP	CC	V	---	PTAYAGKLLISLSEER	---	ISAHVPPNVNATE	C	G	C	R
-----	---	------	-----	--------------	----	---	-----	------------------	-----	---------------	---	---	---	---

TABLE 1 Data collection and phasing statistics

Dataset	Resolution (Å)	Unique reflections (completeness in last resolution range)	R_{merge}^* (%)	Average isomorphous difference (%) (15–3.5 Å)
Native	1.8	14,995 (94.5%, 1.9–1.8 Å)	10.6	—
K ₃ UO ₂ F ₅	2.8	4,123 (84.7%, 2.87–2.8 Å)	7.3†/5.6‡	19.6
UO ₂ (NO ₃) ₂	3.4	2,383 (99.9%, 3.49–3.4 Å)	8.8†/7.8‡	16.7
K ₂ PtCl ₄	2.8	4,194 (98.6%, 2.87–2.8 Å)	8.8	9.0
K ₂ PtCl ₆	2.8	4,161 (95.3%, 2.87–2.8 Å)	10.3	15.0

Dataset	Number of sites	$R_{\text{cullis}}^{\S}$ (%) (15–3.4 Å)	Anomalous data	Phasing power (15–3.4 Å)	Binding site
K ₃ UO ₂ F ₅	1	41.7	yes	2.92	Glu 35
UO ₂ (NO ₃) ₂	1	46.0	yes	2.42	Glu 35
K ₂ PtCl ₄	1	68.4	no	1.73	Asn 69
K ₂ PtCl ₆	1	62.0	no	1.64	Asn 69

Diffraction data were collected using a FAST area detector (Enraf-Nonius, Delft, The Netherlands). Data were evaluated using the program MADNES¹⁸. All other crystallographic calculations were done using the CCP4 program package (Daresbury Laboratory). The space group was determined as $P3_121$ or $P3_221$ by examining the 001 and the hk0 reflections. Heavy-atom derivative sites were determined by inspection of the Harker sections of the difference Patterson maps. Heavy-atom parameters x , y , z and occupancy were refined first, then atomic temperature factors were also refined. The structure was solved using the K₃UO₂F₅ derivative alone with the aid of the anomalous scattering signal of the derivative, which provided SIRAS phases. The initial map at 3.4 Å had a mean figure of merit of 0.596 for 2,343 reflections between 15.0 and 3.4 Å. It was improved by a solvent-flattening procedure¹⁹ (Fig. 3a), estimating a solvent content of 50% and iteration for four cycles. This map allowed the determination of the space group ($P3_221$), the unambiguous chainfold and the location and fitting of 104 of 112 (93%) of the amino-acid side-chains from the known amino-acid sequence²⁰ using the program FRODO²¹. Only the regions 70–71 and 91–96 were unclear before refinement. The heavy-atom site of UO₂²⁺ was found to be near Glu 35, which was reasonable owing to its location on the surface of the molecule and its negative charge. The model had a crystallographic R -factor of 39.4% before refinement, which was done by using a combination of simulated annealing using the program X-PLOR²² and restrained least-squares refinement as implemented in the program TNT²³. Electron density maps calculated with coefficients $2F_o - F_c$ and $F_o - F_c$ at different refinement stages or after omitting parts (typically 8–10 residues) of the structure confirmed the interpretation (Fig. 3b). The model includes 78 water molecules. The atomic temperature factors range from 3 to 99 Å² for protein atoms and 15 to 92 Å² for water molecules. The current R -factor is 18.1% for all reflections between 8.0 and 2.2 Å. The root-mean-square deviation from ideal bond length is 0.01 Å and from bond angles 1.98 degrees. All but one non-glycine residue ϕ , ψ angles lie in the 'allowed' regions of the Ramachandran plot, the exception being residue Asn 42.

* $R_{\text{merge}} = \sum (|I - \langle I \rangle|) / \sum I$, where I is intensity measurement for symmetry-related reflections, and $\langle I \rangle$ is the mean intensity for this reflection.

† Excluding anomalous differences.

‡ Anomalous dispersion signals included.

§ $R_{\text{cullis}} = \sum |F_{\text{PH}} - F_{\text{P}}| / \sum |F_{\text{PH}} - F_{\text{P}}|$, where F_{P} and F_{PH} are the structure factor amplitudes of the protein and the heavy-atom derivative, respectively, and $F_{\text{H(calc)}}$ is the calculated heavy-atom structure factor amplitude summed over centric data only.

|| Phasing power = $(F_{\text{H}})/E$, the r.m.s. heavy-atom structure factor amplitudes divided by the residual lack-of-closure error.

First, the N-terminal segment (Ala 1 to Asp 13) is a region in the sequences of TGF- β 1 to β 5 where many nonconservative amino-acid substitutions are concentrated. The second region of interest is the loop connecting β -strand A and B. The charged and polar residues 25, 26, 31 and 35 are solvent-exposed and residues 23, 25 and 26 show sequence changes between TGF- β 1 to β 5. The third region is the C-terminal end of the long α -helix H₃ (Thr 67 to Ser 73), where the solvent-accessible residue 71 is either Glu or Gly. The fourth region is the most flexible β -turn (Tyr 91 to Pro 96) which connects β -strands C and D₁. Because interactions between a TGF- β molecule and its receptor must be very specific as binding constants are in the picomolar range⁴, it is still unclear whether one or more of these segments interact. The existence of heterodimers like TGF- β 1.2 (refs 1, 16) and TGF- β 2.3 (ref. 16), purified from natural sources, might indicate that one TGF- β -dimer could bind to two different receptors at the same time, as was suggested for receptor types II and III (ref. 17).

So far all isolated or predicted proteins of the TGF superfamily show sequence invariance in the cysteines 15, 44, 48, 77,

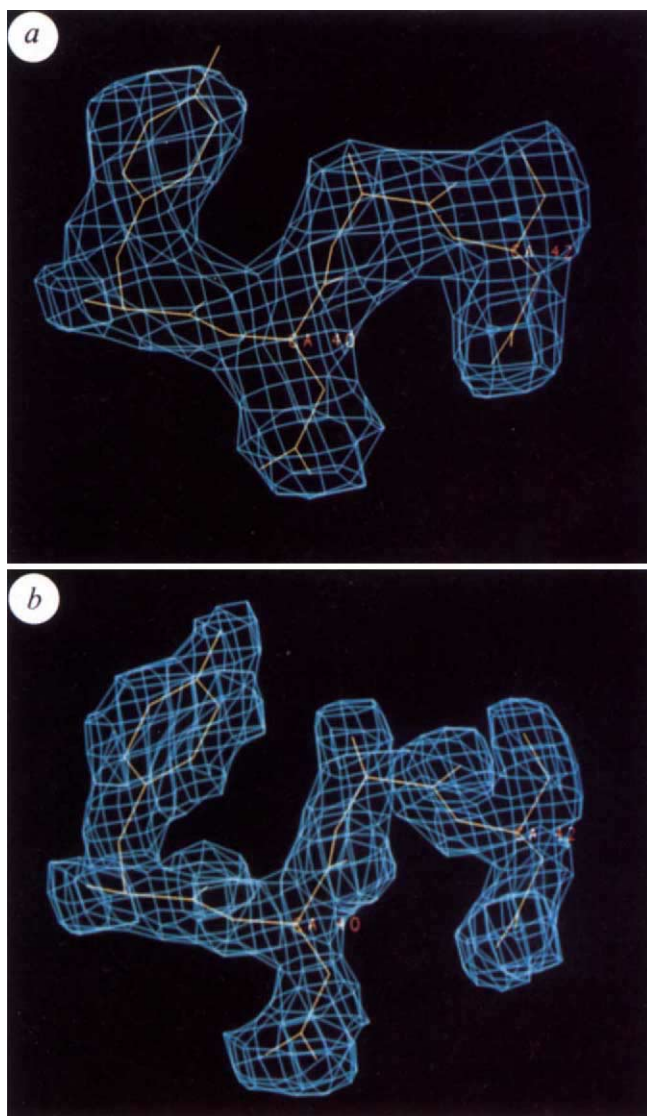


FIG. 3 a, Solvent flattened electron density map at 3.4 Å displaying the region Tyr 39 to Asn 42 calculated with the phases derived from the K₃UO₂F₅ derivative alone and its anomalous signal. b, An equivalent region of the electron density of a $3F_o - 2F_c$ map at 2.2 Å after refinement. Both maps are contoured at 1 r.m.s. above the mean.

78, 109 and 111, as well as Pro 36 and Gly 46 (Fig. 2c; numbering as in TGF- β 2). All the cysteines are involved in the disulphide pattern of TGF- β 2 as well as in the intermolecular disulphide bridge between the monomers. Amino acid 46 in the eight-residue ring has to be a Gly because of steric hindrance and because of the atypical conformation angles ($\phi = 135.1^\circ$, $\psi = 160.9^\circ$). Proline 36 was found to be the only *cis*-Pro in TGF- β 2 and is necessary for ending the long loop and continuing in a β -strand. Many of the amino acids involved in the dimer interface are invariant or show conservative changes in most members of the TGF superfamily. We therefore propose the general fold, including the TGF- β knot, of all the proteins of the TGF superfamily to be the same with variations in the loop regions. The five more closely related members of the TGF subfamily (TGF- β 1 to TGF- β 5) show an overall sequence homology of 64 to 76% (compared to TGF- β 2). The residues involved in the helix-sheet interactions or in the hydrogen-bonding network with the water molecules in the interface are invariant or show conservative changes. Of the 29 amino acids of TGF- β 2 that are involved in the dimer interface (Fig. 2b), 24 are absolutely invariant in all five isoforms. Conservative changes were observed for three positions (residues 43 (Phe to Tyr), 58 (His to Tyr) and 74 (Ala to Ile)), whereas only two amino acids (residues 57 (Gln to Thr) and 68 (Ile to His)) show nonconservative changes. Both variable residues are located in the long α -helix H_3 . This might lead to small positional changes of the α -helix H_3 of one subunit and the β -sheet of the other subunit. All other varying residues are found fairly evenly distributed over the whole surface of the dimer. The general fold of all five TGF- β isoforms, including the TGF- β knot, is therefore likely to be the same and only small changes are expected. The fact that heterodimers TGF- β 1.2 (refs 1, 16), as well as TGF- β 2.3 (ref. 16), have been isolated from natural sources support this proposal.

The first crystal structure of a protein of the TGF superfamily reveals a new fold including a very unusual disulphide pattern, the TGF- β knot. With this information, other three-dimensional structures of TGF isoforms or even of more distantly related proteins could be solved and modelled. Future work must include the analysis of the TGF- β receptor complexes. This will help us understand the biochemical roles of these members of the TGF superfamily and will help in the development of new therapeutic agents, such as for wound healing, bone formation and immune modulation. □

22. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
23. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta Crystallogr.* **A43**, 489–451 (1987).
24. Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
25. Lee, S.-J. *Molec. Endocrin.* **4**, 1034–1040 (1990).
26. Wharton, K. A., Thomson, G. H. & Gelbart, W. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9214–9218 (1991).

ACKNOWLEDGEMENTS. We thank J. P. Priestle for discussions during structure determination and critical review of this manuscript, N. Cerletti, D. A. Cox and G. K. McMaster for providing TGF- β 2 and for discussions, K. von Meyenburg for encouragement and support for this project, M. Carvajal for technical assistance in the crystallization experiments and F. Winkler for bringing the structure of the potato inhibitor of carboxypeptidase A to our attention. The coordinates will be deposited in the Brookhaven Data Bank.

CORRECTIONS

Cloning of cDNAs for Fanconi's anaemia by functional complementation

Craig A. Strathdee, Hanna Gavish, William R. Shannon & Manuel Buchwald

Nature **356**, 763–767 (1992)

WE have discovered that the sequence of our *FACC* gene in Fig. 1c is in error. The corrected sequence appears below, starting at position 787 of the cDNA.

GAG	CGA	GTG	GCG	TCC	CTG	TCA	CGA	GTT
E	R	V	A	S	L	S	R	V
TGT	GTC	CCA	CTT	ATT	ACC	CTG	ACA	
C	V	P	L	I	T	L	T	

The change does not alter any of the conclusions of the paper. The length of the open reading frame remains 557 amino acids. No new significant homologies to the cDNA sequence nor to the translated protein were detected in GenBank, EMBL or Swiss-Prot databases, nor were any new functional motifs detected with the corrected sequence. The correct version has been submitted to the EMBL database. The accession number is X66184. We regret any inconvenience caused by this mistake.

A new type of synthetic peptide library for identifying ligand-binding activity

Kit S. Lam, Sydney E. Salmon, Evan M. Hersh, Victor J. Hruby, Wieslaw M. Kazmeierski & Richard J. Knapp

Nature **354**, 82–84 (1991)

IN this paper we inadvertently omitted to cite the work of Fukura and colleagues (A. Fukura, F. Sebestyen, M. Asgedom and G. Dibo *14th Int. Congr. Biochem.* FR013; 1988), who independently described a similar synthetic method for producing multiple peptide sequences (which we called "split synthesis"). However, Fukura *et al.* did not describe the concept of 'one bead, one peptide' which was central to our approach.

Received 14 May; accepted 17 July 1992.

1. Cheifetz, S. *et al. Cell* **48**, 409–415 (1987).
2. Roberts, A. B. & Sporn, M. B. *Adv. Cancer Res.* **51**, 107–145 (1988).
3. Massagué, J. A. *Rev. Cell Biol.* **6**, 597–641 (1990).
4. Cheifetz, S. & Massagué, J. *J. Biol. Chem.* **266**, 20767–20772 (1991).
5. Lin, H. Y., Wang, X.-F., Ng-Eaton, E., Weinberg, R. A. & Lodish, H. F. *Cell* **68**, 775–785 (1992).
6. Cox, D. A., Kunz, S., Cerletti, N., McMaster, G. K. & Bürk, R. R. in *Angiogenesis: Key Principles, Science, Technology, Medicine* (eds Steiner, R., Weisz, B. & Langer, R.) 287–295 (Birkhäuser, Basel, Switzerland, 1992).
7. Joyce, M. E., Roberts, A. B., Sporn, M. B. & Bolander, M. E. *J. Cell Biol.* **110**, 2195–2207 (1990).
8. Wrann, M. *et al. EMBO J.* **6**, 1633–1636 (1987).
9. Mason, A. J. *et al. Nature* **318**, 659–663 (1985).
10. Wozney, J. M. *et al. Science* **242**, 1528–1534 (1988).
11. Wozney, J. M. *Prog. Growth Factor Res.* **1**, 267–280 (1989).
12. Schlunegger, M. P. *et al. FEBS Lett.* **303**, 91–93 (1992).
13. Tamaoki, H. *et al. Prot. Engng.* **4**, 509–518 (1991).
14. Rees, D. C. & Lipscomb, W. N. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4633–4637 (1980).
15. Janin, J. & Wodak, S. *J. molec. Biol.* **125**, 357–386 (1978).
16. Ogawa, Y. *et al. J. Biol. Chem.* **267**, 2325–2328 (1992).
17. Wang, X.-F. *et al. Cell* **67**, 797–805 (1991).
18. Messerschmidt, A. & Pflügrath, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
19. Wang, B.-C. *Meth. Enzym.* **115**, 90–112 (1985).
20. Ikeda, T., Lioubin, M. N. & Marquardt, H. *Biochemistry* **26**, 2406–2410 (1987).
21. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1987).

Measuring up in the marketplace

Featured this week — an ultra-low-noise patch-clamp amplifier, a 120-channel signal-stacking seismograph and alpha/beta liquid scintillation counters that feature automatic crossover optimization and ultra-low-level counting.

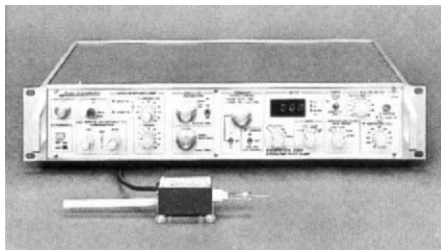
THE new AccuRATE 1000 intelligent **gas flowmeter** from J & W Scientific handles flow rates from 0.1 ml min^{-1} to $1,000 \text{ ml min}^{-1}$ and is designed to replace 'bubble-type' flow-rate measurement methods (*Reader Service No. 100*). Even split-flow ratios are intended to be simple. J & W says the \$495 (US) AccuRATE 1000 flowmeter is accurate to within ± 3 per cent and can be used with all noncorrosive gases. A self-diagnostic program runs each time AccuRATE 1000 is turned on. If the instrument is not used for a period of 10 min, the power automatically trips off, conserving the 9-V battery.

An analog bandwidth of 100 MHz, a sampling rate on each channel of 20 megasamples per second and a maximum memory length of 32,000 words per channel are features of the two-channel Yokogawa DL1100 **digital oscilloscope** from Martron Instruments (*Reader Service No. 101*). This compact model features a separate analog/digital converter on each channel and has a large, 7-inch CRT display with alphanumeric and cursor readout facilities. Other features include waveform processing and analysis functions and an optional memory-card reader for programming settings or saving recorded data. The high sampling rate allows a bandwidth of up to 8 MHz to be achieved for single-shot waveform measurements and the large memory allows the pre/post trigger point to be set over a wide range for the capture of random events. The DL1000 range of oscilloscopes incorporates several digital techniques that are designed to emulate the feel of an analog oscilloscope. High-speed processing circuitry, for example, gives a screen refresh rate of 60 Hz, thus providing the effect of a real-time display.

■ Patch clamps and pico-injectors

The PLI-100 **pico-injector** from Medical Systems can be used to microinject volumes ranging from microlitres to femtolitres through micropipettes and is intended for biomedical applications at the cellular level (*Reader Service No. 102*). The company says that reproducible delivery is achieved by precise regulation of pressure and time. Vacuum is applied initially to fill the delivery pipette. Regulated balance pressure is applied between injections to prevent initial dilution of the delivered material by capillary action.

Like its predecessor, the Axopatch 200, the Axopatch 200A integrating **patch clamp** from Axon Instruments is designed for



Capacitor-feedback patch clamp.

single-channel, bilayer and whole-cell voltage- and current-clamp recording (*Reader Service No. 103*). However, the design of the capacitive-feedback circuit in the new Axopatch 200A headstage allows single-channel recording at exceptionally low noise. The open-circuit noise in the \$6,900 (US) Axopatch 200A is 0.16 pA rms in a 10-kHz bandwidth (8-pole Bessel filter). The headstage switches between capacitor-feedback for single-channel recording and resistor-feedback for whole-cell recording. Other enhancements in the Axopatch 200A include the addition of ZAP (a convenient measure that allows electronic rupture of the patch), dual-speed current clamp (to allow faster current clamping in small cells), the ability to use the holding command to set current commands in current clamp and a choice of two gain settings on the dedicated current output.

■ Benchtop Instrumentation

The portable, 120-channel 9000-A **signal-stacking seismograph** from Bison Instruments supersedes the company's earlier 9000 series model (*Reader Service No. 104*). It offers one to 120 channels that are keyboard-selectable with onboard stacker and sample intervals as fast as 50 μs . The 120-channel system allows record lengths of 500–12,000 samples (500 sample steps) and a frequency response of 4–4,000 Hz.



Bison's signal-stacking seismograph.

High-pass and low-pass analog filters are standard. The 9000-A is equipped with a high-resolution thermal plotter and RS232, IEEE and SCSI ports for computer control and tape interface. The instrument requires a 12 V d.c. power supply (160 W).

Designed primarily for the biotechnology market, the Kompact Maldi **benchtop mass spectrometers** from Kratos Analytical not only offer biochemists a rapid method of determining accurate molecular weights but also offer the ability to differentiate between closely related peptides (*Reader Service No. 105*). Two models are available: the Kompact Maldi-II and the Kompact Maldi-III, which are priced at \$99,950 and \$149,950 (US), respectively. The two instruments are fully automatic and utilize laser-desorption time-of-flight (TOF) techniques. Whereas both configurations feature linear TOF technology, the compact Maldi-III includes a reflectron TOF analyser. The biochemist prepares samples onto a disposable sample holder using a dedicated sample preparation device and then, unlike conventional wait-and-queue regimes, visits the Kompact Maldi on a multi-user, walk-up-and-use basis. The sample holder is then placed in the instrument for immediate analysis. The instruments feature sample holders with a capacity of up to 20 samples, automatic sample indexing and fast data processing. All procedures are computer-controlled by a 32-bit UNIX workstation. Extensive data processing capabilities are provided, including peptide calculator and predictor tools, which interpret and give indications about proposed sequences from acquired biopolymer analysis data.

Researchers making measurements of electrophoretic mobilities, zeta potentials and sub-micrometre particle sizes may be interested in the new Zetasizer 4 **particle characterization instrument** from Malvern Instruments (*Reader Service No. 106*). The Zetasizer 4 combines classical microelectrophoresis techniques with Malvern's underlying strengths in photon correlation spectroscopy. Malvern says revised sample handling and new sample cell designs simplify changeovers between high and low conductivity media. The instrument is capable of measurements with both positive and negative particles. Particle size analysis, including true multi-angle operation, covers the range 3 nm to 3 μm . A 70-station autosampler/diluter and multi-tasking software provide a high degree of automation.

ADVERTISEMENTS

\$3.50/Base

**Custom DNA
Synthesis**

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

Reader Service No.33

WOLFE PEPTIDES

30 mg 20 mer
not purified

80,- DM per
residue

TIB
MOLBIOL

for further information :
TIB MOLBIOL Berlin
FAX +49 30 712 7734

Reader Service No.17

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.30 per base, no setup!

(0.2 μ mole - limited offer)

option of H.P.L.C. purification - £30

biotinylation, fluorescent labelling

phosphothioates, many other modifications

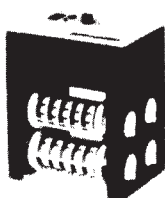
OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No. 9

BIANORM
EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify
the binding parameters and
thermodynamic data between
ligands and biopolymers.
Results come closest to
the actual values.
Sample volume from
0.2 ml to 5.0 ml.
Up to 20 experiments
in parallel.

BIANORM
Gerte
P.O. Box 650126
D-5000 Munich 65,
FRG

Reader Service No.45

Packard's new line of Tri-Carb 2500TR/AB and 2550TR/AB **alpha/beta liquid scintillation analysers** feature automatic, quantitative, crossover optimization and ultra-low-level counting (*Reader Service No. 107*). These systems integrate advanced

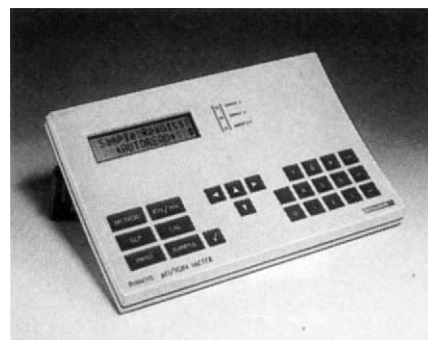


Alpha/beta liquid scintillation systems.

pulse decay analysis with time-resolved liquid scintillation counting to provide simultaneous gross alpha/beta and low-level beta counting in a multipurpose instrument. The Tri-Carb TR/AB systems automatically optimize alpha and beta separation into simultaneous, high-resolution, multichannel analysers. Alpha, beta and overlapped alpha/beta spectra are displayed in real-time on the colour CRT monitor. Crossover is automatically minimized for each sample type and optimum settings are stored for automatic recall. Quantitative crossover curves are displayed and plotted upon request, enabling the operator to fine-tune the system to suit special sample requirements. These network-compatible systems can automatically access resident application software for complete unattended counting and data processing for up to 30 different users or applications. The TR/AB systems offer a sample capacity of 408 large vials, 720 small vials or a combination of both sizes.

■ Metering devices

The PHM95 **pH/ion meter** from Radiometer is designed to be equally suitable for high-precision research analyses and routine measurements of mV, pH and ion concentration (*Reader Service No. 108*). For optimal flexibility, 16 methods are provided, which can be edited individually. Each method contains all the information necessary for an application, including type of electrode, calibration procedure, result acceptance criteria and number of digits in the result. The method link facility and the



Radiometer's all-in-one pH/ion meter.

two electrode inputs enable two different measurements to be performed in the same sample. Users can choose between three different pH calibration procedures: 'auto' with recognition of four buffers, 'fixed' with selection of 13 preprogrammed buffers, or with 'free' entry of buffer values. The calibration curve can be defined and stored with up to nine calibration points for ion concentration measurements. Standard addition/subtraction and analate addition/subtraction techniques are provided for ion concentration measurements.

According to World Precision Instruments, its Dri-Ref **reference electrode** offers stability, low resistance and very low electrolyte leakage (*Reader Service No. 109*). It is designed for measurements that demand minimal sample contamination by the electrode. Dri-Ref's internal electrode is made with a large-surface silver/silver chloride pellet, which, the company says, has been proven to be stable, long lasting and very low in resistance. Total resistance is less than 500 Ω — an important consideration when making low-noise measurements. Although the internal filling solution contains KCl, WPI says that low fluid leakage means that Dri-Ref can be used in combination with ion-selective electrodes, including those for K^+ and Cl^- , without contamination by the reference electrode. The epoxy body of the Dri-Ref electrode is designed to be chemically resistant to strong acids, alkalis and organic solvents.

■ Liquid dispensers

Hamilton announces the introduction of its latest **automated pipettor/diluter**, the Microlab 500, which combines transfer pipetting, automated diluting and dispensing into one instrument (*Reader Service No. 110*). The Microlab 500 can function as a single- or dual-syringe pipettor, diluter or dispenser, and uses a simple user interface with an easy-to-read LCD and touch-sensitive membrane keyboard. The Microlab 500 can accommodate volumes ranging from 1 μ l to 25 ml.

The new Multidrop **programmable microplate dispenser** from Labsystems is capable of filling a plate (20 μ l per well) in less than 5 s (*Reader Service No. 111*). Users can select dispensing volumes from 20 to 390 μ l in 10- μ l increments. In addition, the extended reservoirs allow dispensing of up to eight different reagents simultaneously. According to Labsystems, even viscous solutions such as agar can be dispensed. The dispensing head can be detached and autoclaved for aseptic procedures. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

Growth in jobs with a green tinge

Sandra MacPherson

Increased awareness of the importance of the environment, accompanied by tighter legislation, has increased the number of jobs available in conservation, but has also made them highly prized.

THE sheer diversity of positions available should offer some hope to anyone determined to break into conservation work. Qualified people for specialized jobs are still in demand, and there is an increasing need for interdisciplinary scientific expertise, but competition continues to increase.

Several types of organization employ staff in a whole gamut of environment-related positions — from the day-to-day running of reserves through environmental audit to basic research. Governments are still the principal employers, both nationally and locally, though non-governmental and campaigning organizations are important, and often high profile, both locally and internationally. Private industry is increasingly employing its own conservation or environment staff, and specialist consultancy firms are springing up to fill any gaps.

Government

In the UK last year responsibility for the environment was split between three organizations which report to the Department of the Environment: English Nature, Scottish Natural Heritage and the Countryside Council for Wales.

The department itself employs very few staff with conservation remits, a mere three wildlife inspectors based in Bristol. English Nature is the largest employer with more than 700 staff. Jobs vary from site managers on national nature reserves, through conservation officers carrying out surveys and environmental assessments, to scientific specialists collating research and commissioning projects.

The outcome of research can influence legislation and lead to the designation of new areas as Sites of Special Scientific Interest. Competition for positions is fierce. A recent advertisement for 30 vacancies brought in 3,000 applications. Relevant experience, enthusiasm and a first- or second-class degree in an environmentally based subject are crucial.

Pay is based on civil service scales. Scientific officers start on £10,551 and can work up to senior posts on £21,797. Site managers — who need a minimum of five years' experience — start on £17,925 and work up to £22,941 at senior grade. The Joint Nature Conservation Committee receives funds from each of the three agencies but decides its own research remit. It has 50 full-time conservation staff and 30 on short-term

contracts. They research and advise on specific issues such as marine life or lower vascular plants.

Another body funded by the UK government is the Countryside Commission. It employs 230 staff, but owns no land and manages no facilities. Countryside officers, who start on £13,000, assess grant proposals for projects such as land reclamation schemes and check to see that the work is completed. A recent advertisement for about 10 short-term posts attracted 2,500 enquiries.

Most other western European countries, notably France, have devolved re-

than those in natural resources. For example, the Environmental Protection Agency, which has jurisdiction over all pollution laws, is set to recruit 1,500 staff over the next five years. But the number of staff in the National Parks Service, which manages 361 parks, has remained stable over the past 10 years.

The number employed in the parks service is still quite high, however. There are 14,000 full-time employees and 8,000 seasonal employees. A scientific researcher with academic qualifications and experience will start on a salary of \$30,000, rising to \$45,000 after ten years. Such a person has to thrive on a measure of public controversy — decisions made by the service are not always popular.

Outside government

Most of the large number of conservation groups worldwide are small and rely heavily on volunteers. Others are international networks with a core staff and teams of consultants.

In the United States, which is estimated to have 3.5 million conservation and environment jobs, there are more than 10,000 different, private, non-profit environmental groups.

In eastern Europe conservation is still very rudimentary because there is no money to train or employ conservation professionals. But voluntary groups are growing in number and improving rapidly, and the World Bank and the EC are both financing large projects.

Small conservation groups are thriving in the UK too. The Royal Society for Nature Conservation, the umbrella organization for the UK's 47 local wildlife trusts and 44 urban conservation groups, estimates that the number of jobs in wildlife and habitat management has risen by a quarter in the past five years. Together, these groups employ 700 full-time staff, but only 210 of those are involved in hands-on conservation. The trusts rely heavily on their increasingly knowledgeable volunteer resource to do the groundwork — bad news for conservationists hoping for paid employment.

Conservation officers earn between £8,000 and £15,000 a year, but job satisfaction is said to compensate for the low wages. There is no shortage of interest — posts attract up to 200 applicants. The Devon Wildlife Trust, for example, has grown over the past three and a half years from seven employees, 3,500 members and an annual turnover



sponsibility for the environment to local authorities. In Germany it is the *Länder* which take responsibility. This devolution tends to produce a more generalist approach, with staff unable to devote their full time to conservation. And there is less likely to be expertise on tap when needed. Local governments in the UK are increasingly employing their own conservation staff. Many jobs have been created by directives from the European Commission (EC), which, for example, require all local authorities to have a recycling policy by August this year.

In the United States, local governments do employ conservation staff, but there are also 16 government agencies which deal with the environment. Chief among employers of conservation and research staff are the Environmental Protection Agency, the National Parks Service, the Forest Service, the Bureau of Land Management and the Fish and Wildlife Service — the guiding agency for the Endangered Species Act.

Environment jobs are increasing faster

of £70,000 to 31 full-time staff, 6,500 members and a turnover of £500,000. Such jobs are primarily concerned with establishing a database of knowledge about the local resource. Though positions are limited, a spokesman for the Devon Wildlife Trust said that most local groups are happy to employ people with ideas for fundraising to finance their own employment costs.

The Royal Society for the Protection of Birds is a non-governmental organization (NGO) on a grand scale. During the past three years, membership has risen from 560,000 to 840,000, permanent staff by 140 to 600, and temporary staff by 90 to 330. Staff with direct conservation remits include 9 working to acquire nature reserves, 12 permanent and 12 temporary ecologists, 254 wardens, 100 seasonal reserve staff, 33 general research staff, 80 contract staff on specific research projects, 30 campaigning staff and 7 investigators who travel the world posing as wild bird dealers to catch smugglers. The latter posts, three of which are held by ex-police officers, are particularly popular.

International

Large international NGOs still initiate many of the projects in the developing world, though countries such as Venezuela, Kenya and India are starting to take more responsibility for their own environment.

The International Union for the Conservation of Nature and Natural Resources (IUCN) coordinates projects on national parks and protected areas, and species and habitat survival. It employs about 50 full-time scientific staff at its Geneva headquarters and 40 worldwide. Local staff are preferred. IUCN also has a register of consultants, listing 1,700 individuals who are usually recognized experts in their field.

IUCN founded the World Conservation Monitoring Centre in Cambridge, UK, which collects global data on species, protected areas and wildlife trade. It employs 35 staff with scientific backgrounds and computer literacy who input and analyse data collected by contract field staff.

High-profile campaigning jobs are notorious for attracting overqualified people who want to combine science with policy. In the early days of campaigning organizations, the only prerequisite was great enthusiasm for the cause. But with the advent of increasingly specialized campaigns, and a drive for professionalism, a scientific degree has become almost a necessity.

Greenpeace, for example, employs scientists both directly and on a consultancy basis, usually within their own countries. Salary is determined by the local cost of living. At Greenpeace UK,

many of the campaigners have science backgrounds, but there are only three permanent staff in the science division itself. Greenpeace International employs about 1,000 staff in 38 offices worldwide.

Greenpeace International also has a laboratory at Exeter University where four scientists do research work used to back campaigns. One of the team described the work as 'intellectually mind boggling' — very stimulating, very difficult and very rewarding. Accuracy is essential as all published results will be scrupulously analysed by those trying to pick holes in Greenpeace campaigns.

Friends of the Earth, which has 47 groups around the world, is also increasingly employing scientists in its specialized campaign teams. A senior campaigner can expect a salary of up to £21,000. An assistant campaigner starts on about £14,000. Friends of the Earth has no field research posts. Local consultants are used when necessary.

The World Wide Fund for Nature, on the other hand, employs scientists primarily in an advisory capacity. The 20 conservation staff in the UK head office oversee international projects. The organization seems to be moving away from employing biologists and zoologists in favour of staff with a background in natural resources or agriculture.

The United Nations Environment Programme, based in Kenya, is the environmental conscience of the UN. Its purpose is to promote environmental action worldwide by winning the cooperation of governments, international scientific and professional communities and NGOs. It has only 200 staff, and relies heavily on consultants to do the research which feeds the large data networks established to help countries make environmental management decisions.

Private Industry

The increasing onus on private industry to take responsibility for its waste products and their environmental impact means that more companies are finding it worthwhile to employ their own environmental staff. Others use consultancies, but that too creates jobs.

This need within industry has not gone unnoticed. The group environmental laboratory at ICI in the UK, for example, receives about 2,000 unsolicited CVs each year. Advertised jobs attract an additional 200 to 1,000 applicants each.

The laboratory now employs about 90 staff looking at ways to manage effluent. The team, which has doubled in size over the past five years, advises on pollution control, environmental management and the implications of new processes for the local environments. Studies may involve, for example, computer models of water flow in an estuary. In a more directly commercial role, the

ICI agrochemicals group employs some 100 scientists to investigate the potential environmental effects of pesticides.

Considering all these types of environment-related employment, ICI estimates that there are 20 UK job opportunities (to fill new and existing environment-related posts) each year, and several times that number worldwide. Graduates can expect to start on £13,000, or £17,000 with a PhD.

Shell UK is another example. It has an environmental affairs department responsible for environmental impact assessments, keeping tabs on emissions and generally improving environmental awareness. The department has grown from six staff four years ago to eight staff now, but is not expected to grow further.

Private consultancies

Two years ago private environmental consultancies were expanding rapidly on the back of the boom in development and the wave of new legislation. Demand for environmental impact assessments in the UK has, however, slowed with the recession.

The situation is not quite as bad in other parts of Europe, particularly in the south, where EC funds to decrease regional disparity are being given to major infrastructure projects, such as road building. Such large projects require extensive impact assessment and are attracting many UK and US consultancies.

A broad background tends to be useful in the typical small consultancy — a first degree in biology or ecology followed by an MSc in pollution control, for example. Junior consultants with two to three years' experience can expect a salary of about £17,000. Project leaders can command in excess of £30,000.

Many architectural firms, engineering practices and construction companies have set up their own environmental consultancy services, and universities, too, are beginning to sell their skills on a consultancy basis. Lancaster University's Institute of Environmental Sciences, for example, now does a considerable amount of contract work.

At Liverpool University, consultancy has gone one step further. The environmental advisory unit, set up within the university, has become a separate company. There are now 54 staff, 44 of whom are consultants. Five years ago there were only 25. Two posts were advertised recently: an environmental auditor and a terrestrial ecologist. Hundreds of applications were received, but the post of environmental auditor is being readvertised because none of the applicants had the appropriate skills.

Sandra MacPherson is a science writer for General Practitioner with an interest in the environment.